



UNIVERSITÉ DE NANTES



**IAE NANTES**  
ÉCONOMIE & MANAGEMENT

Master 1 Économétrie et Statistiques, parcours Économétrie Appliquée

# **Étude sur le taux de pauvreté en 2015**

Analyse économétrique sur 59 pays à travers le monde

## **Auteurs :**

Nicolas DAVOINE

Diane THIERRY

**Année universitaire 2019-2020**

# Sommaire :

Partie 1 : Analyse économique -----	2
Partie 2 : Analyse économétrique-----	23
Partie 3 : Conclusion et discussion des résultats -----	50
Partie 4 : Table des matières -----	53
Partie 5 : Annexes -----	55
Partie 6 : Bibliographie -----	63

# **PARTIE 1 : Analyse économique**

## **I- Abstract**

Dear future president of Tadjikistan, 2 months ago you asked us to make an analysis of the poverty rate in the world in 2015. This study aim to understand the main determinants of poverty, to ultimately trying to eradicate it in your country. This is the main argument of your presidential campaign, so we worked hardly during 2 months to give you the key of your future success.

Therefore, in this paper we study potential determinants of poverty rate in 59 countries around the world in 2015. We take into account 9 differents variables that we suspect to have an individual impact on the poverty rate level of a country. In order to determine the correlation between the poverty rate and all the explanatory variables, we first built a model composed by 3 exogenous and non-collinear variables to be able to use the MCO method. We chose unemployment rate, health spendings, and the Gini index. However, the existence of heteroscedastic residues made use change the Gini index variable for the corruption index, nevertheless, the outcome was the same.

As a consequence we decided to use the Two-Stage Least Squares (2SLS) method, with unemployment rate, health spendings and average wage. We choose Gini index and R&D spendings as instrumental variables for the medium wage variable. The results showed that the R-squared adjusted was too low (0,35). As a consequence we took the decision to switch the variable medium wage for the Human Development Index. Through tests we noticed that, the HDI wasn't endogenous but exogenous, and so we use the MCO method to analyse the poverty rate variations. Through the results we saw that health spendings aren't reliable to reflect the health of a population, and that HDI is relevant to influence the poverty rate.

## II- Introduction

Dans cette introduction nous allons vous présenter l'évolution historique de la pauvreté dans le monde et ses différents aspects afin de contextualiser notre étude en comprenant les origines et le fonctionnement de la pauvreté, il sera plus facile pour vous de la contrer dans votre pays, vous pourrez aussi en parler plus aisément lors des campagnes présidentielles qui se dérouleront en janvier 2020.

La pauvreté est un fléau mondial principalement alimenté par les crises économiques, financières mais aussi historiques ou encore climatiques. Les scientifiques, les décideurs politiques et les professionnels du développement s'accordent sur l'aspect «multidimensionnel» de la pauvreté. Avant la première moitié du XXème siècle, la grande majorité de la population mondiale vivait dans la pauvreté, seuls les pays d'Europe de l'Ouest et d'Amérique du Nord étaient suffisamment développés pour répondre aux besoins primaires de la majeure partie de leur population. Ce n'est qu'à la fin de la seconde Guerre Mondiale que la plupart des pays de la planète coopérèrent afin d'éradiquer à long terme la pauvreté dans le monde.

Influencées par l'expérience du plan Marshall Américain (1948-1952) et par la signature des accords de Bretton Woods en 1944, de nombreuses institutions financières internationales ont été créées telles que la Banque Mondiale, le Fond Monétaire International (FMI) ou encore l'Association Internationale de Développement (IDA). Parmi les objectifs de ces institutions, le plus important est celui d'aider les pays considérés comme 'pauvres' à amorcer leur développement économique pour réduire la misère dans ces pays, afin de maintenir une paix durable à l'échelle mondiale. En effet, l'Aide Publique au Développement (APD) s'établit dès les années 1960 comme un outil de coopération internationale.

De ce fait ces organismes ont accordé des prêts à long terme aux pays les plus pauvres de la planète. Cela a permis de financer des programmes aidants à (re)construire les politiques, les institutions, les infrastructures et le capital humain nécessaires à un

développement à la fois équitable et soutenable. Ces interventions ont contribué à faire baisser la pauvreté et l'extrême pauvreté dans le monde<sup>1</sup>.

Depuis les années 1950, le nombre de ménages considérés comme pauvres a diminué également grâce aux vagues successives d'immigration de populations des pays du Sud allant vers le Nord pour bénéficier des meilleures conditions de vie. L'**immigration** désigne l'entrée de personnes étrangères dans un pays ou une aire géographique pour y séjourner quelque temps ou pour s'y installer. Ainsi, les ménages fuyant la misère de leur pays d'origine ont pu avoir accès aux conditions de vie des pays d'accueils (généralement les Etats-Unis et les pays d'Europe de l'Ouest). De ce fait, le nombre de personnes vivant dans la pauvreté s'est réduit dans les pays du Sud, sans forcément augmenter dans les pays développés puisque leurs économies de marchés permettaient d'employer aussi les nouveaux entrants.

Les répercussions économiques d'une telle vague d'immigration sont fortes, en effet les remises correspondant transferts financiers de l'épargne des migrants à leurs proches (familles, amis, communauté) restés dans le pays d'origine, deviennent une source importante de financement et de développement. Selon la Banque Mondiale, en 2010, le montant de ces envois d'argent s'élevait à 239 milliards de dollars. Temporairement ralenti par la crise financière et économique de 2008, le volume des remises n'a cependant cessé de croître depuis le début des années 2000<sup>2</sup>.

Ainsi, le rapport sur les inégalités mondiales 2018 montre que depuis 1980, les populations les plus pauvres du monde (c'est à dire les 50% du bas de la distribution des revenus) ont vu leurs revenus augmenter de manière significative, grâce à l'essor des pays émergents<sup>3</sup>.

Aujourd'hui il n'y a jamais eu aussi peu de personnes vivant dans la pauvreté si l'on rapporte ce chiffre à la population mondiale. Néanmoins on constate depuis plusieurs décennies que les Etats des pays les plus avancés ont du mal à éradiquer la misère au sein

---

<sup>1</sup><http://www.sciencespo.fr/ceri/fr/carto/dossier-carto-la-pauvrete-dans-le-monde> (consulté le 24/09)

<sup>2</sup><http://www.sciencespo.fr/ceri/fr/carto/dossier-carto-la-pauvrete-dans-le-monde> (consulté le 24/09)

<sup>3</sup><http://ses.ens-lyon.fr/actualites/rapports-etudes-et-4-pages/rapport-sur-les-inegalites-mondiales-2018-wid-world-decembre-2017> (consulté le 26/09)

même de leur pays : c'est le cas des pays les plus riches en termes de PIB dans le monde (Etats-Unis, Chine, Japon...).

De plus, bien que les écarts de pauvreté entre les pays développés et les pays en développement se soient considérablement réduits au cours des cinquante dernières années, ils demeurent importants et restent une préoccupation phare des dirigeants politiques. Il est donc primordial de savoir quels facteurs contribuent à réduire ou accroître la pauvreté dans un pays pour pouvoir expliquer rationnellement les écarts de niveau de pauvreté entre les pays. Pour ce faire nous étudierons le taux de pauvreté en 2015, en appliquant les différentes méthodes d'estimation sur un échantillon de 59 pays à travers le monde.

Avant cela, nous tenons à préciser que notre projet ne s'inspire d'aucun ouvrage ou étude déjà réalisés. Contrairement aux ouvrages et études publiées précédemment sur le sujet, nous ne ferons aucune séparation ou regroupement en fonction des types d'individus touchés par la pauvreté. C'est en effet très souvent le cas dans les études réalisées ; elles distinguent et analysent séparément des catégories d'individus afin d'effectuer des comparaisons entre celles-ci. Par exemple, on se penche séparément sur la pauvreté chez les jeunes et chez les plus âgés, ou encore sur les actifs occupés et les chômeurs, dans le but de déterminer des solutions adaptées à chaque catégorie.

Dans notre analyse nous étudierons la pauvreté en termes de quantité d'individus sans aucune distinction, et à l'échelle globale d'un pays. Le but de notre démarche est de pouvoir faire des comparaisons simples entre les différents pays qui permettent de mettre en lumière les impacts individuels qu'ont les différentes variables explicatives sur le taux de pauvreté global d'un pays.

Nous allons à présent vous présenter notre base de données élaborée spécialement pour votre campagne, et justifier chacune des variables utilisées pour ce dossier.

### III- Analyse économique des variables retenues

Le modèle de régression linéaire multiple de nos variables est le suivant :

$$Tx\_pauv_i = \alpha + \beta_1 * X_{1i} + \beta_2 * X_{2i} + \beta_3 * X_{3i} + \beta_4 * X_{4i} + \beta_5 * X_{5i} + \beta_6 * X_{6i} + \beta_7 * X_{7i} + \beta_8 * X_{8i} + \beta_9 * X_{9i} + u_i$$

Où :

- $Y_i$  : “Tx\_pauv” est la variable à expliquer, c’est celle que nous allons analyser en utilisant 9 variables explicatives. Elle représente le taux de pauvreté en 2015, exprimé en pourcentage de la population
- $X_1$  : “PIB\_hab” est le PIB par habitant exprimé en dollars par habitant (\$/hab)
- $X_2$  : “Sal\_moy” est le salaire moyen exprimé en dollars par habitant (\$/hab)
- $X_3$  : “Gini” est l’indice de Gini qui représente les inégalités de distribution du revenu dans un pays “i”
- $X_4$  : “Corrup” est l’indice de corruption : plus il est proche de 0, plus grande est la corruption dans le pays concerné
- $X_5$  : “IDH” est l’indice de développement humain compris entre 0 (situation catastrophique) et 100% (parfait développement humain)
- $X_6$  : “Tx\_chom” est le taux de chômage exprimé en pourcentage de la population
- $X_7$  : “Sante” sont les dépenses en Santé réalisées par le gouvernement, en pourcentage de son PIB
- $X_8$  : “Espce\_vie” est l’espérance de vie à la naissance exprimée en années
- $X_9$  : “RetD” sont les dépenses intérieures brutes de recherche et développement exprimées en pourcentage du PIB

Et :

- i est le nombre de pays compris dans l’échantillon  $\rightarrow i=1,2,3,\dots,59$
- $\alpha, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8, \beta_9$  sont les paramètres du modèle
- $u_i$  est le terme d’erreur associé au modèle théorique

## **A) Une variable à expliquer**

Comme vous le savez, la variable que nous tenterons d'expliquer dans notre analyse est le taux de pauvreté dans près de 59 pays du monde en 2015.

Comme l'explique l'Insee, "le taux de pauvreté correspond à la proportion d'individus dont le niveau de vie est inférieur pour une année donnée à un seuil, dénommé "seuil de pauvreté".<sup>4</sup> Le niveau de vie désigne le revenu disponible, c'est à dire les revenus d'activité (nets des cotisations sociales), les revenus du patrimoine, les transferts en provenance d'autres ménages et les prestations sociales (y compris les pensions de retraite et les indemnités de chômage), nets des impôts directs.<sup>5</sup>

Ainsi, on dit d'un individu qu'il est pauvre s'il vit dans un ménage dont le niveau de vie est inférieur au seuil de pauvreté. Ce dernier peut être défini de manière absolue (en fonction d'un panier de consommation minimale) ou relative (en pourcentage du revenu médian ou moyen). Le choix du seuil pris en compte dans l'analyse de la pauvreté est très important, puisque cette analyse peut amener à des résultats radicalement différents selon l'option de calcul choisie.<sup>6</sup> L'Insee, tout comme l'Eurostat et les autres pays européens, mesure la pauvreté monétaire de manière relative alors que d'autres pays (comme les États-Unis ou le Canada) ont une approche absolue (fondée sur l'évaluation de la dépense minimale nécessaire pour assurer un niveau de vie "tolérable").

Dans ce dossier nous allons retenir le seuil dit "relatif". Cette méthode de calcul tient compte du niveau de vie d'un pays : la pauvreté étudiée est fonction du niveau de revenu médian de la population.<sup>7</sup> Dans les données que nous avons récoltées sur le taux de pauvreté sur les sites officiels de l'OCDE et de la Banque Mondiale, le seuil fixé s'élevait à 50% du revenu médian, cela signifie qu'est considéré comme pauvre tout individu ayant un revenu disponible inférieur à la moitié du revenu médian.

---

<sup>4</sup> <https://www.insee.fr/fr/metadonnees/definition/c1320> (consulté le 29/09)

<sup>5</sup> <https://www.insee.fr/fr/metadonnees/definition/c1458> (consulté le 29/09)

<sup>6</sup> [https://fr.wikipedia.org/wiki/Seuil\\_de\\_pauvret%C3%A9](https://fr.wikipedia.org/wiki/Seuil_de_pauvret%C3%A9) (consulté le 30/09)

<sup>7</sup> [https://fr.wikipedia.org/wiki/Seuil\\_de\\_pauvret%C3%A9](https://fr.wikipedia.org/wiki/Seuil_de_pauvret%C3%A9) (consulté le 30/09)



De même, la pauvreté peut être reflétée au travers de l'évolution de la distribution des revenus. Il est toutefois très difficile de définir et de mesurer la "pauvreté" en soi, en particulier lorsque l'on procède à des comparaisons internationales. En effet, les besoins de la population, ainsi que les réseaux sociaux et familiaux, ne sont pas les mêmes d'un pays à l'autre, et certains services publics gratuits dans un pays peuvent être payants ailleurs. Ainsi, un niveau de revenu donné peut se traduire par un niveau de vie différent selon les pays<sup>8</sup>. Par manque de données ou tout simplement par le fait que ces subtilités de la pauvreté sont difficilement quantifiables et comparables entre les pays, nous ne les inclurons pas dans notre analyse.

Nous essayerons cependant de donner un aperçu le plus précis possible de la pauvreté dans les 59 pays concernés, avec toutes les données dont nous avons pu disposer ; nous chercherons à modéliser de manière optimale ses principaux déterminants.

## **B) Et 9 variables explicatives**

Afin d'expliquer les niveaux de taux de pauvreté dans les pays observés nous allons étudier des variables qui, en théorie, ont une incidence sur la pauvreté au niveau global d'un pays. Certaines de ces variables font référence à la pauvreté monétaire des individus (comme le PIB par habitant et le salaire moyen), d'autres sont davantage liées à la qualité de vie et à la santé (comme l'IDH, l'espérance de vie, et la part des dépenses de santé dans le PIB). Finalement d'autres sont liées plus indirectement à la pauvreté des individus d'un pays (comme l'indice de corruption, le coefficient de Gini, le taux de chômage et la part de la recherche et développement dans le PIB).

Les données que nous avons récoltées proviennent essentiellement des bases de données de l'OCDE<sup>9</sup> et de la banque mondiale<sup>10</sup>. À partir de ces informations nous avons créé notre propre base de données<sup>11</sup>, avec laquelle nous avons pu réaliser notre analyse (graphiques élaborés nous même) et tirer nos conclusions.

---

<sup>8</sup> <https://www.cairn.info/revue-economique-de-l-ocde-2002-1-page-7.htm> (consulté le 23/10)

<sup>9</sup> <https://data.oecd.org/fr/> (consulté le 22/10)

<sup>10</sup> <https://donnees.banquemondiale.org/> (consulté le 22/10)

<sup>11</sup> [https://docs.google.com/spreadsheets/d/e/2PACX-1vRs8IjziBAAeFhiOgClMbc6iMVn88lh14LjPz4JxAteINLOxnAy748VRHGuu\\_vBft1IiMID\\_ARo1K6i/pub?output=xlsx](https://docs.google.com/spreadsheets/d/e/2PACX-1vRs8IjziBAAeFhiOgClMbc6iMVn88lh14LjPz4JxAteINLOxnAy748VRHGuu_vBft1IiMID_ARo1K6i/pub?output=xlsx)

Pour chaque variable explicative nous avons élaboré un graphique de corrélation avec la variable à expliquer pour pouvoir justifier de manière visuelle le choix de nos variables :

➤ une corrélation positive, où la droite de régression serait croissante, signifie que lorsque la variable explicative étudiée augmente, la pauvreté fera de même.

➤ une corrélation négative, où la droite de régression serait décroissante, signifie que lorsque la variable explicative étudiée augmente, la pauvreté quant à elle diminuera.

De plus, nous avons trouvé pertinent de distinguer nos observations par continent pour rendre compte des disparités de la pauvreté entre ceux-ci.

Nous voulions intégrer dans nos variables explicatives, la part de la population des plus de 25 ans ayant un diplôme de l'enseignement supérieur comme un master ou équivalent. Cette variable est intéressante puisqu'elle prend en compte l'impact de l'éducation sur la pauvreté d'un pays. Depuis les années 1970, l'accès à l'éducation est une condition nécessaire pour avoir un travail avec une rémunération suffisamment élevée pour sortir de la pauvreté monétaire. Aujourd'hui, non seulement cette éducation est devenue absolument indispensable mais en plus elle ne garantit plus automatiquement l'accès à un travail bien rémunéré : dans les pays développés comme en développement, il est nécessaire d'avoir un diplôme de l'enseignement supérieur pour bénéficier d'une rémunération convenable. Nous n'avons cependant pas trouvé assez de données pour l'année 2015, c'est pourquoi nous intégrerons pas cette variable à notre analyse.

Le meilleur moyen d'appréhender le niveau de pauvreté consiste à s'intéresser à deux facteurs : les écarts de revenus marchands et l'impact de la fiscalité et de la politique sociale sur la redistribution. Nous aurions aimé nous focaliser sur les effets des impôts et transferts, c'est-à-dire sur les sommes que les pouvoirs publics prélèvent sur un groupe pour en faire bénéficier un autre, en ajoutant une variable à expliquer. Néanmoins les données mesurant les différents niveaux de redistribution entre les pays étaient insuffisantes, c'est pourquoi nous ne les incluons pas non plus dans notre analyse.

**→ Le problème d'endogénéité de certaines variables**

Une des hypothèses fondamentales permettant l'utilisation des Moindres Carrés Ordinaires (MCO) est l'exogénéité des variables explicatives. Cette hypothèse suppose que les erreurs du modèle sont indépendantes des variables explicatives.

Nous soupçonnons les variables "PIB\_hab", "Sal\_moyen" et "Espce\_vie" d'avoir un effet *feedback* sur le taux de pauvreté : le taux de pauvreté dans un pays influence son PIB par habitant et son salaire moyen. En effet, plus une population est pauvre moins, elle a de ressources financières pour consommer, ce qui freine l'activité économique du pays et entraîne la baisse du PIB par habitant et des salaires moyens. De même, la pauvreté affecte l'espérance de vie des individus : de nombreuses enquêtes démontrent que les personnes dites "pauvres" sont davantage exposées aux risques de santé et vivent moins longtemps.<sup>12</sup> Les pauvres ont moins accès que les autres individus aux soins de base : en moyenne, d'après une étude réalisée par la Banque Mondiale dans 56 pays en développement, l'accès aux soins de santé primaire augmente fortement avec le statut socio-économique du ménage mesuré par son revenu.<sup>13</sup>

Cela veut dire que ces trois variables explicatives ne sont pas exogènes mais endogènes. Nous ne pourrions donc pas utiliser les MCO si nous intégrons de telles variables dans notre régression, puisque cela viole l'hypothèse d'exogénéité des variables.

### 1- Pib par habitant ( $X_1$ ) et salaire moyen ( $X_2$ )

Les variables  $X_1$  et  $X_2$  sont liées puisqu'il s'agit du PIB par habitant et du salaire moyen des pays concernés. Ces deux variables visent à donner un aperçu du niveau de revenu moyen dont disposent les individus des différents pays, qui par association devraient nous donner un aperçu approximatif de la pauvreté des pays.

→ Le **produit intérieur brut par habitant** correspond à la valeur du PIB divisée par le nombre d'habitants d'un pays. Il est donc plus efficace que le PIB pour mesurer le

---

<sup>12</sup> <https://www.insee.fr/fr/statistiques/3319895> (consulté le 22/10)

<sup>13</sup> <http://www.cafepedagogique.net/lemensuel/lenseignant/schumaines/ses/Pages/2008/Sant%C3%A9etPauvre%C3%A9danslespaysend%C3%A9veloppement.aspx> (consulté le 22/10)

développement d'un pays, cependant, il n'est qu'une moyenne ne permettant pas de rendre compte des inégalités de revenu et de richesse au sein d'une population.<sup>14</sup>

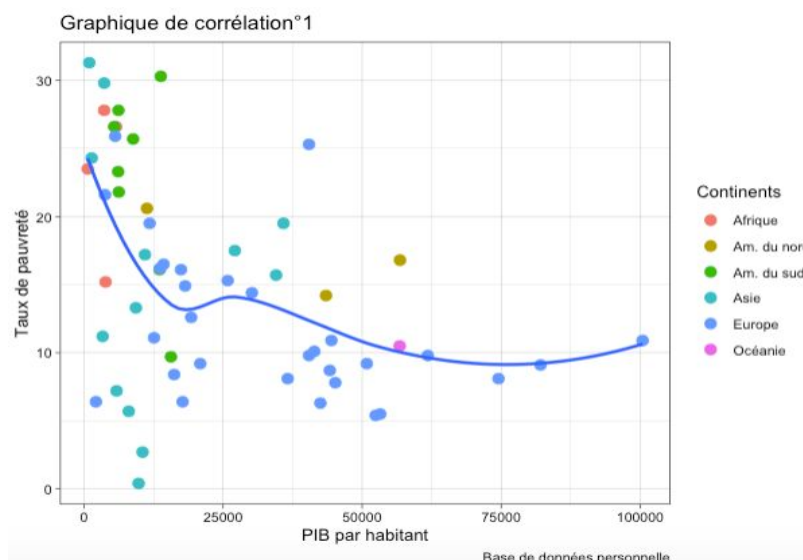
Selon l'office fédéral de la statistique Suisse, la croissance du PIB par habitant dépend de deux facteurs principaux liés au travail des individus :

➤ la productivité du travail (c'est à dire "*le rapport entre le PIB et le nombre d'heures effectivement travaillées sur le territoire économique considéré*"<sup>15</sup>).

➤ le nombre moyen d'heures effectives de travail par habitant sur une année.

Les données sur le PIB par habitant des pays pris en compte dans notre analyse sont toutes converties en dollars US, de telle sorte que le pouvoir d'achat réel des différentes devises puisse être corrigé grâce à la Parité de Pouvoir d'Achat (PPA).

### GRAPHIQUE N°1 : Relation entre le PIB par habitant et le taux de pauvreté



Grâce au graphique n°1 nous remarquons une corrélation négative assez intuitive : lorsque le PIB par habitant - c'est à dire la richesse détenue par un individu, augmente, alors le taux de pauvreté diminue dans le pays concerné. Il est intéressant de noter les différences de pauvreté et de PIB par habitant en fonction du continent, les pays d'Afrique, d'Amérique du Sud et certains d'Asie ont un taux de pauvreté relativement élevé pour un PIB par habitant plutôt faible. A contrario, les pays ayant un PIB par habitant élevé et un taux de pauvreté bas sont principalement des pays d'Europe et d'Océanie (il s'agit de l'Australie).

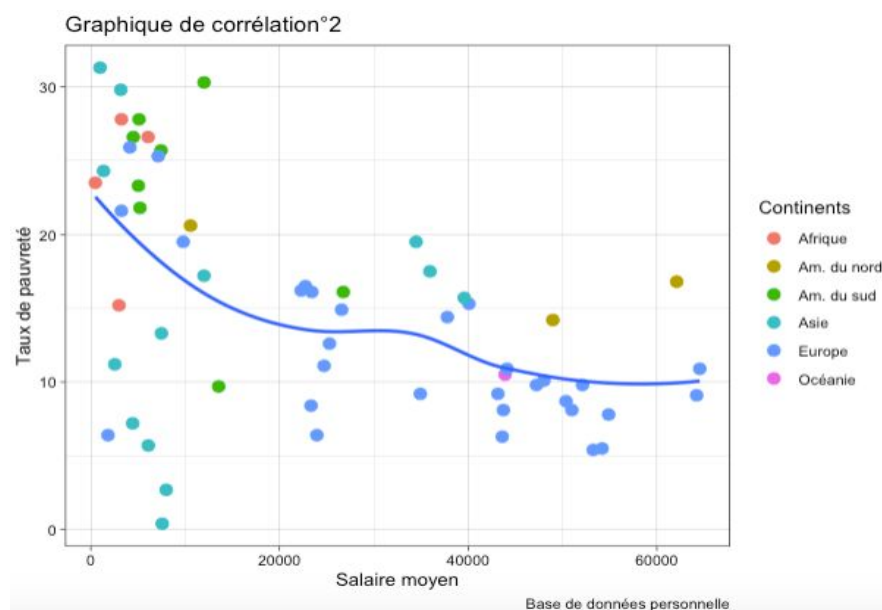
<sup>14</sup> [https://fr.wikipedia.org/wiki/Produit\\_int%C3%A9rieur\\_brut\\_par\\_habitant](https://fr.wikipedia.org/wiki/Produit_int%C3%A9rieur_brut_par_habitant) (consulté le 08/10)

<sup>15</sup> <https://www.ifrap.org/etat-et-collectivites/pib-par-habitant-quelle-tendance-en-europe-et-en-france> (consulté le 09/10)

→ La seconde variable explicative est le niveau de **salaire moyen** dans le pays. Il est obtenu en divisant la masse salariale totale issue des comptes nationaux par le nombre moyen de salariés dans l'ensemble de l'économie, puis en multipliant le chiffre obtenu par le nombre d'heures hebdomadaires moyennes habituelles pour un salarié à temps plein, rapporté à la moyenne des heures hebdomadaires travaillées pour tous les salariés. Cet indicateur est mesuré en prix constants en USD.<sup>16</sup> Contrairement au PIB par habitant, le salaire moyen nous donne exclusivement une indication sur la rémunération du travail sans prendre en compte les autres formes de revenu dont les individus peuvent bénéficier comme les revenus fonciers et les revenus du capital.

Cet indicateur est néanmoins utile pour mesurer la pauvreté d'un pays puisqu'il permet de comparer de manière approximative les pays en termes de travailleurs pauvres. En effet, plus le salaire moyen du pays est faible, plus il serait logique qu'il y ait des travailleurs pauvres, d'autant plus que depuis les années 2000, le nombre de travailleurs pauvres augmente dans presque tous les pays développés.

GRAPHIQUE N°2 : Relation entre le salaire moyen et le taux de pauvreté



Comme pour la variable « PIB/habitant », on peut noter ici une corrélation négative avec notre variable à expliquer. Cela valide l'hypothèse selon laquelle PIB par habitant et

<sup>16</sup> <https://data.oecd.org/fr/earnwage/salaires-moyens.htm> (consulté le 15/10)

salaires moyens sont liés, ils donnent tous deux une approximation du niveau de vie moyen d'un pays.

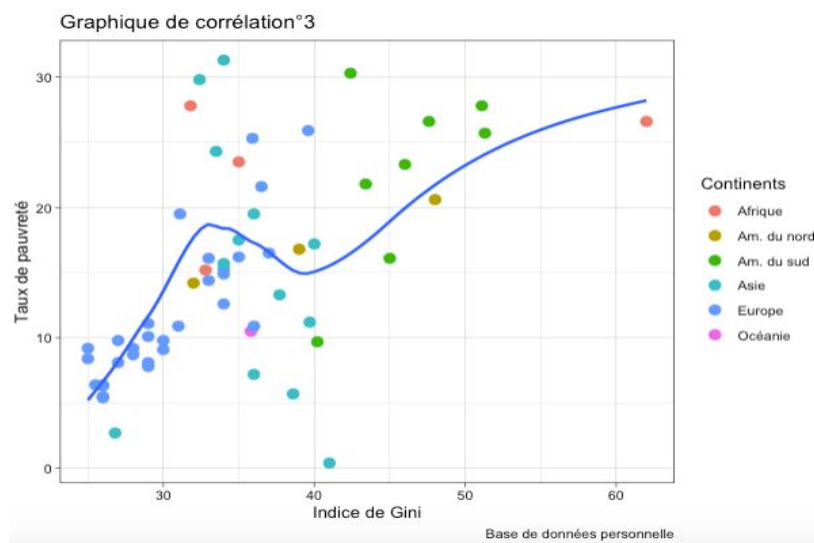
## 2- Indice de Gini ( $X_3$ )

Les inégalités de revenu sont mesurées ici grâce au **coefficient de Gini**. Ce dernier, variant de 0 et 1, indique la répartition de la richesse entre les individus d'un territoire donné. Plus il est proche de zéro, plus la répartition des revenus est égalitaire, à contrario plus il s'approche de 1 plus la richesse est concentrée sur un petit groupe de personnes.<sup>17</sup>

Nous avons décidé d'utiliser cet indicateur comme variable explicative du taux de pauvreté parce que la pauvreté dans un pays est fortement liée aux inégalités de richesse. En effet plus la richesse est concentrée sur certains individus, moins le quartile le plus pauvre de la population aura de richesse.

Il existe trois formes d'inégalités : inégalités monétaires, inégalités des conditions de vie et inégalités des chances. Dans notre analyse nous nous intéresserons exclusivement à l'inégalité monétaire. Cette dernière pose la question de la répartition des fruits de la croissance économique entre les différentes tranches de la population.

**GRAPHIQUE N°3 : Relation entre l'Indice de Gini et le taux de pauvreté**



D'après le graphique n°3 nous remarquons une relation positive assez nette entre Y et l'indice de Gini. Lorsque l'indice est proche de 0 le taux de pauvreté est lui aussi proche de 0 puisque cela signifie que le revenu est réparti également sur le territoire ; d'où la

<sup>17</sup> <https://data.oecd.org/fr/inequality/inegalite-de-revenu.htm> (consulté le 13/10)

relation positive (courbe croissante). Nous notons que les pays développés se situent davantage en bas à gauche du graphique où les inégalités ainsi que le taux de pauvreté sont les plus faibles. En haut de la droite de régression nous voyons que les continents s'y trouvant sont : Afrique et Amérique du Sud.

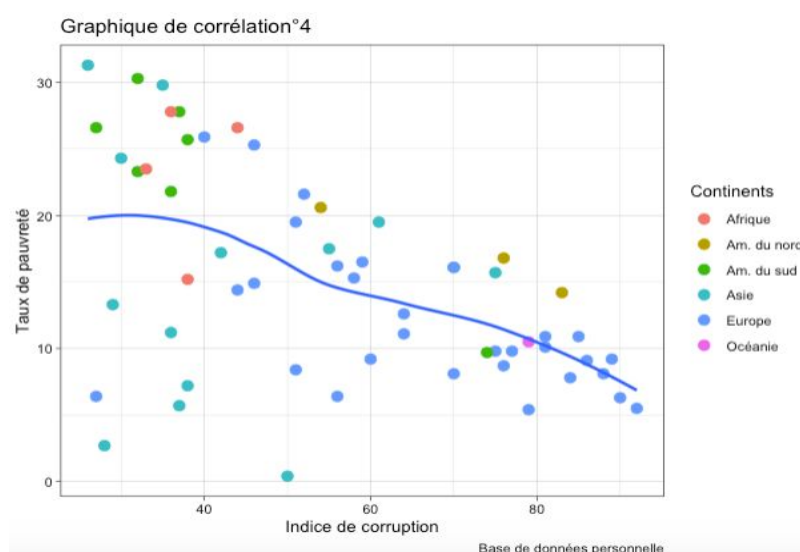
### 3- Indice de corruption (X<sub>4</sub>)

L'**indice de corruption** est élaboré à l'aide d'enquêtes réalisées auprès d'hommes d'affaires, d'analystes de risques et d'universitaires résidant dans un pays ou à l'étranger.<sup>18</sup> Il sera proche de 0 si le niveau de corruption est élevé, et proche de 1 s'il y a peu ou pas de corruption. La corruption, l'inégalité et la pauvreté forment un cercle vicieux auto-entretenu par la distribution inégale du pouvoir dans la société et la répartition inégale de la richesse.

De fait, la corruption perpétue la pauvreté, elle biaise les décisions et détourne des fonds déjà peu nombreux, en refusant aux populations pauvres l'accès aux services sociaux de base ainsi qu'aux ressources qui leur permettraient d'améliorer leurs conditions de vie.<sup>19</sup>

Par conséquent, la corruption accroît la vulnérabilité des ménages à la pauvreté en cas de perte de revenus, et empêche les ménages les plus pauvres d'en sortir, puisque l'Etat ne peut pas venir en aide de manière efficace à travers les transferts sociaux comme il le devrait.

GRAPHIQUE N°4 : Relation entre l'Indice de corruption et le taux de pauvreté



<sup>18</sup> [https://fr.wikipedia.org/wiki/Indice\\_de\\_perception\\_de\\_la\\_corruption](https://fr.wikipedia.org/wiki/Indice_de_perception_de_la_corruption) (consulté le 16/10)

<sup>19</sup> [https://www.transparency.org/files/content/activity/2012\\_PCA\\_brochure\\_FR.pdf](https://www.transparency.org/files/content/activity/2012_PCA_brochure_FR.pdf) (consulté le 16/10)

On constate qu'il existe une relation négative entre le taux de pauvreté et l'indice de corruption, ce qui confirme notre hypothèse d'interdépendance négative entre ces deux variables.

#### **4- Indice de développement humain (X<sub>5</sub>)**

L'**IDH** est une mesure créée par le Programme des Nations Unies pour le Développement (PNUD) qui permet de mieux mesurer les inégalités mondiales et de trouver des solutions aux problèmes d'éducation et de pauvreté. L'indice de développement humain se mesure à partir de trois critères principaux : le produit intérieur brut (PIB) par habitant, l'espérance de vie des citoyens d'un Etat et le niveau d'éducation mesuré à partir de 15 ans et plus. En faisant entrer l'éducation et l'espérance de vie de la population dans sa grille de lecture, cet indice de mesure permet d'être plus précis dans l'analyse du développement des Etats et de leur population.

Comme expliqué ci-dessus, l>IDH a pour objectif de mesurer le niveau de développement des pays. Cependant il ne s'arrête pas au simple poids économique des pays ou à leur niveau de vie, mais il intègre des données plus qualitatives telles que l'éducation par exemple.<sup>20</sup> Cette variable est donc particulièrement intéressante puisqu'elle inclut une dimension éducative que nous n'avions pas pu intégrer dans notre analyse à cause du manque de données disponibles sur la variable "part de la population de plus de 25 ans possédant un diplôme de l'enseignement supérieur".

L>IDH se présente comme un nombre compris entre 0 et 1 : plus l'indice est proche de 1, plus le niveau de développement du pays est élevé. Bien que celui-ci ait considérablement progressé dans toutes les régions du monde entre 1990 et 2015, une personne sur trois vit toujours dans de faibles conditions de développement humain.<sup>21</sup>

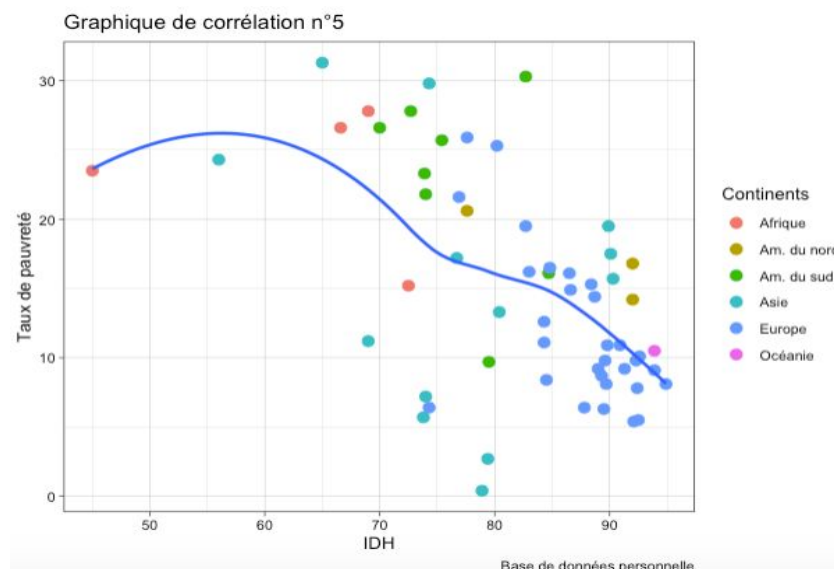
#### **GRAPHIQUE N°5 : Relation entre l>IDH et le taux de pauvreté**

---

<sup>20</sup> <http://ses.webclass.fr/notion/indice-developpement-humain-idh> (consulté le 17/10)

<sup>21</sup> <https://www.undp.org/content/undp/fr/home/librarypage/hdr/> (consulté le 15/10)





Grâce au graphique n°5 on remarque une corrélation négative entre Y et la variable IDH : au delà de 65-70% de niveau de développement, le taux de pauvreté a tendance à être plus faible dans les pays étudiés. On note aussi une certaine concentration des pays en terme d'IDH en fonction du taux de pauvreté, la grande majorité des pays se situent dans la partie droite du graphique, où l'IDH est compris entre 70 et 95%.

## 5- Taux de chômage (X<sub>6</sub>)

Selon le Bureau International du Travail (BIT), un chômeur est une personne en âge de travailler qui répond simultanément à trois conditions :

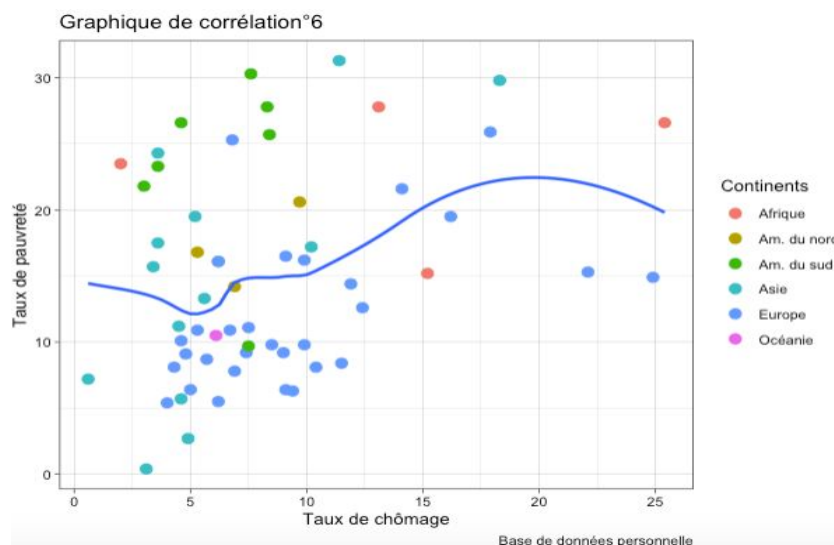
- être sans emploi, c'est à dire ne pas avoir travaillé au moins une heure durant une semaine de référence
- être disponible pour prendre un emploi dans les 15 jours
- avoir cherché activement un emploi dans le mois précédent ou en avoir trouvé un qui commence dans moins de trois mois.<sup>22</sup>

Le **taux de chômage** correspond donc à la part de la population active étant au chômage et cherchant activement un emploi ; ils sont considérés comme des “actifs inoccupés”. Il est aisé de faire le lien entre la pauvreté et l'inactivité. Lorsqu'un individu est inactif il ne perçoit pas de revenus d'activité, alors que pour la majeure partie de la population, leurs ressources financières proviennent essentiellement du revenu lié au travail (surtout pour les individus les plus pauvres).

<sup>22</sup> <https://www.insee.fr/fr/metadonnees/definition/c1129> (consulté le 17/10)

La question serait alors de savoir comment cette exposition à des périodes d'inactivités se répercute sur le risque de pauvreté. De plus, dans la plupart des pays du monde les transferts sociaux sont insuffisants pour permettre aux individus de vivre décemment et d'échapper à la pauvreté, c'est pour cela que les actifs occupés, notamment ceux qui occupent un emploi toute l'année à temps complet, apparaissent toujours beaucoup moins exposés à la pauvreté monétaire que les chômeurs.<sup>23</sup>

GRAPHIQUE N°6 : Relation entre le taux de chômage et le taux de pauvreté



On voit grâce au graphique n°6, qu'il existe une corrélation positive entre le taux de chômage et celui de pauvreté. Il est intéressant de noter qu'avec une échelle presque similaire pour les 2 axes, la hausse du taux de pauvreté rapporté au taux de chômage est plus que proportionnelle : le taux de pauvreté augmente plus rapidement que le taux de chômage.

## 6- Dépenses de Santé ( $X_7$ ) et espérance de vie à la naissance ( $X_8$ )

→ Tout d'abord, la variable explicative  $X_7$  correspond à la part du PIB consacrée aux **dépenses de santé**, elle est donc exprimée en pourcentage du PIB. Indirectement elle donne une indication sur la part de la richesse que les individus allouent à se soigner pour être en meilleure santé. Cependant, nous considérons cette variable comme insuffisante pour tenter de mesurer l'impacte que l'état de santé d'une population a sur la pauvreté. Cela est notamment dû au fait que les frais médicaux sont très variables d'un pays à un autre, par

<sup>23</sup> [http://www.onpes.gouv.fr/IMG/pdf/travaux2005-2006-1-2-2-chomage\\_remunerations\\_pauvrete.pdf](http://www.onpes.gouv.fr/IMG/pdf/travaux2005-2006-1-2-2-chomage_remunerations_pauvrete.pdf)  
(consulté le 19/10)

exemple aux Etats-Unis ils sont très élevés alors qu'au Canada beaucoup moins, c'est aussi lié au fait que les disparités en termes de PIB peuvent biaiser les comparaisons faites entre les pays.

→ C'est pourquoi la variable explicative  $X_8$  qui correspond à l'**espérance de vie** à la naissance, peut être un bon complément à la variable dépenses de santé pour avoir un aperçu de l'état de santé de la population puisqu'un individu 'x' en bonne santé a tendance à vivre plus longtemps qu'un individu 'y' en mauvaise santé. Ainsi, une espérance de vie élevée dans un pays nous informe que sa population vit généralement en bonne santé. A contrario, dans un pays où l'espérance de vie serait faible cela signifierait que les conditions de vie et/ou de santé y sont médiocres, ce qui implique une pauvreté plus élevée dans le pays concerné <sup>24</sup>.

Au vu de cette relation cause à effet, nous attendons donc une corrélation négative entre le taux de pauvreté et l'espérance de vie à la naissance, ce que nous vérifierons juste après quelques explications sur l'importance de la santé, et son lien intuitif avec le niveau de pauvreté.

#### ⇒ Explications du lien entre la santé (dépenses en santé et espérance de vie) et la pauvreté

Le lien qui existe entre l'état de santé de la population et le taux de pauvreté est subtil et comporte plusieurs approches. Tout d'abord lorsque la majeure partie de la population n'a pas accès aux soins les plus basiques ou refuse de dépenser une partie de leur revenu afin de remédier à des soucis de santé, cela affecte leur capacité à travailler. Les individus sont généralement moins productifs et plus fréquemment en arrêt maladie ce qui réduit leurs revenus salariaux car ils subissent une perte de revenu très importante en cas d'inactivité.

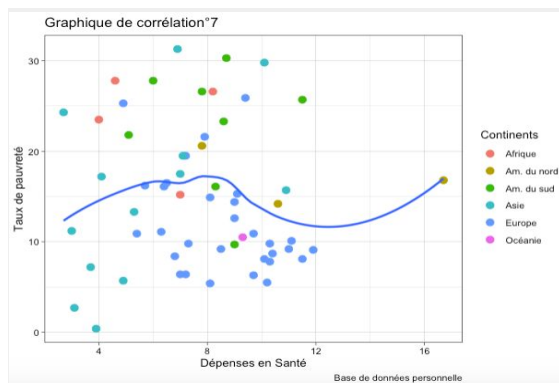
De plus, les performances scolaires des enfants sont liées à leur état de santé. Or si les enfants d'un pays ne peuvent pas effectuer leurs études dans le meilleur état de santé possible, ils risquent de faire des études plus courtes et d'avoir de moins bons résultats que les enfants des autres pays, ce qui les pénalisera à terme lorsqu'ils seront sur le marché du travail. En effet, il apparaît qu'ils seront moins enclins à occuper un travail à hautes qualifications et seront obligés d'accepter des emplois moins qualifiés donc moins bien rémunérés, ils seront alors plus sensibles au risque de chômage. On peut donc conclure que les pauvres sont en partie pauvres parce qu'ils sont en mauvaise santé.

---

<sup>24</sup> <https://www.insee.fr/statistiques/fichier> (consulté le 20/10)

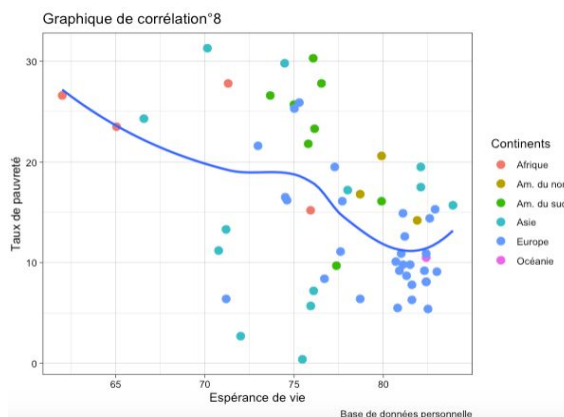
Bien que les variables explicatives  $X_7$  et  $X_8$  que nous allons utiliser dans cette analyse ne mesurent pas directement le niveau de santé d'une population, elles restent pertinentes pour en donner un aperçu fiable dans beaucoup de pays.

#### GRAPHIQUE N°7 : Relation entre les dépenses en santé et le taux de pauvreté



Il est plus difficile ici d'identifier la nature de la corrélation entre taux de pauvreté et dépenses de santé en % du PIB. Entre 0 et 12% du PIB dédié à la santé des individus, on retrouve des pays avec des taux de pauvreté variant entre 0 et 30%. La corrélation négative attendue n'apparaît pas assez clairement pour valider notre hypothèse formulée précédemment.

#### GRAPHIQUE N°8 : Relation entre l'espérance de vie et le taux de pauvreté

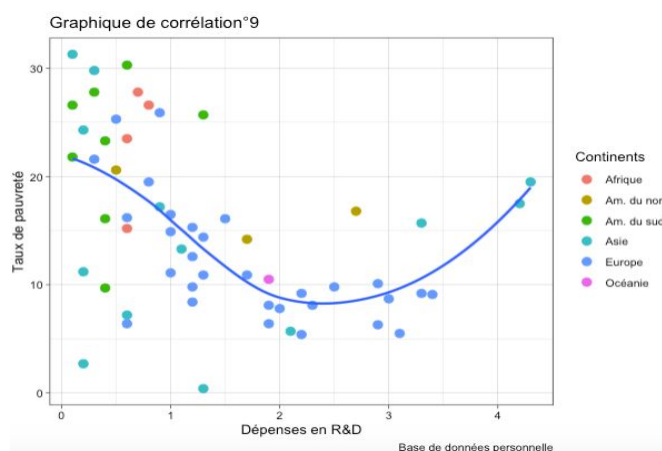


Ici encore on remarque une relation négative entre le taux de pauvreté et l'espérance de vie à la naissance, ce qui confirme nos intuitions sur la corrélation existante entre ces deux variables. On constate les mêmes disparités entre les continents : l'espérance de vie est élevée (+ 80 ans) pour les pays d'Europe principalement, d'Océanie et d'Asie. Alors qu'elle est plutôt faible (- 75 ans) pour les pays d'Afrique, d'autres d'Asie et d'Amérique du Sud.

## 7- Dépenses intérieures brutes de recherche et développement (X<sub>9</sub>)

Les dépenses intérieures de recherche et développement, aussi appelées **DIRD** correspondent, selon l'INSEE "aux travaux de recherche et développement (R&D) exécutés sur le territoire national quelle que soit l'origine des fonds. Une partie est exécutée par les administrations, l'autre par les entreprises".<sup>25</sup> Cet indicateur regroupe donc les dépenses publiques comme privées et est exprimé en pourcentage du PIB. Il existe trois types d'activités au sein des DIRD : la recherche fondamentale (travaux de recherche théoriques), la recherche appliquée (discerner les applications possibles du type n°1) et le développement expérimental (travaux systématiques réalisés pour lancer la fabrication).<sup>26</sup> Les travaux de recherche et développement financés par les dépenses en R&D peuvent avoir de fortes répercussions sur le niveau de vie d'un pays : ils permettent d'accroître les connaissances dans de nombreux domaines d'application tels que la science, les technologies mais aussi la culture, l'humain et la société.

GRAPHIQUE N°9 : Relation entre les dépenses R&D et le taux de pauvreté



Finalement, d'après le graphique de corrélation n°9, nous avons pu vérifier l'hypothèse émise précédemment : il existe bien une relation négative entre les dépenses intérieures brutes de R&D et le taux de pauvreté. Cependant nous observons que la droite de régression a tendance à "se redresser" à la fin - cela est principalement dû aux 2 pays d'Asie qui semblent être atypiques.

<sup>25</sup> <https://www.insee.fr/fr/metadonnees/definition/c1226> (consulté le 28/10)

<sup>26</sup> <https://www.enseignementsup-recherche.gouv.fr/reperes/telechar/res/res00/rapmeto00.pdf> (consulté le 28/10)

⇒ Pour terminer cette partie explicative de nos variables nous allons présenter un tableau récapitulatif reprenant, pour chaque continent, les caractéristiques visuellement observables pour chacune des variables selon nos graphiques de corrélation.

TABLEAU N°1 : Récapitulatif des corrélations entre les différents X et Y

	<b>Afrique</b> (4 pays)	<b>Am.Nord</b> (3 pays)	<b>Am.Sud</b> (8 pays)	<b>Asie</b> (13 pays)	<b>Europe</b> (30 pays)	<b>Océanie*</b> (1 pays)
<b>PIB par habitant</b>	très faible	faible et modérée	plutôt faible	très faible et modérée	répartition homogène (faible à élevée)	élevée
<b>Salaire moyen</b>	très faible	faible et élevée	très faible et modérée	très faible et modérée	modérée et élevée	élevée
<b>Indice de Gini</b>	modérée et très élevée	modérée	élevée	faible et modérée	modérée	plutôt faible
<b>Indice de corruption</b>	faible	modérée et élevée	faible	faible et modérée	modérée et élevée	élevée
<b>IDH</b>	très faible et faible	modérée et élevée	faible et modérée	faible, modérée, élevée	modérée et élevée	très élevée
<b>Taux de chômage</b>	faible, modérée et élevée	faible	faible	faible et modérée	faible, modérée et élevée	faible
<b>Espérance de vie</b>	très faible et modérée	élevée	modérée	très faible, faible et élevée	modérée et élevée	élevée
<b>Dépenses Santé</b>	faible et modérée	modérée et très élevée	modérée	faibles, modérée et élevée	modérée et élevée	modérée
<b>Dépenses R&amp;D</b>	faible	faible et modérée	faible	faible et modérée	faible, modérée et élevée	modérée

\* Notre base de données comporte seulement 1 pays faisant partie de l'Océanie : l'Australie. Bien que celui-ci représente 90% du continent total en termes de superficie, l'échantillon reste peu représentatif donc nos conclusions seront très approximatives.

Au vu des résultats du tableau nous pouvons classer les continents en 4 principaux groupes :

- Pauvre/défavorisé : Afrique
- En voie de développement : Amérique du Sud
- Aisés : Asie, Amérique du Nord
- Riches/développés : Europe et Océanie

On remarque donc de fortes disparités au sein de notre échantillon. De plus il semble y avoir une grande hétérogénéité dans les pays d'Asie comme d'Europe ; pour la plupart des variables on retrouve des valeurs étant faibles, moyennes, ou élevées au sein d'une même variable. Voici pour finir un tableau récapitulatif de nos variables (mesure, nature et source).

TABLEAU N°2 : Tableau de description des données

	<b>Nom</b>	<b>Mesure</b>	<b>Nature</b>	<b>Source</b>
<b>Taux de pauvreté</b>	Y	% de la pop	Quant. continue	OCDE et BM
<b>PIB par habitant</b>	X <sub>1</sub>	\$ / habitant	Quant. continue	OCDE et BM
<b>Salaire moyen</b>	X <sub>2</sub>	\$ / habitant	Quant. continue	OCDE et BM
<b>Indice de Gini</b>	X <sub>3</sub>	0<X <sub>3</sub> <100%	Quant. continue	TI
<b>Indice de corruption</b>	X <sub>4</sub>	0<X <sub>4</sub> <100%	Quant. continue	UNDP
<b>IDH</b>	X <sub>5</sub>	0<X <sub>5</sub> <100%	Quant. continue	OCDE et BM
<b>Taux de chômage</b>	X <sub>6</sub>	% de la pop	Quant. continue	OCDE et BM
<b>Espérance de vie</b>	X <sub>7</sub>	années	Quant. discrète	OCDE et BM
<b>Dépenses Santé</b>	X <sub>8</sub>	% du PIB	Quant. continue	OCDE et BM
<b>Dépenses R&amp;D</b>	X <sub>9</sub>	% du PIB	Quant. continue	OCDE et BM

OCDE : Organisation de coopération et de développement économiques, BM : Banque Mondiale, TI : Transparency International, UNDP : United Nations Development Programme.

## **PARTIE 2 : Analyse économétrique**

Le principe même de l'économétrie consiste à étudier des phénomènes économiques à partir d'observations statistiques, afin de conceptualiser des relations entre les variables pour expliquer au mieux ces phénomènes. C'est ce que nous allons effectuer dans ce dossier ; nous étudierons l'impact qu'ont eu les 9 variables explicatives choisies sur le taux de pauvreté en 2015 dans les 59 pays concernés.

### **I- Nettoyage de la base de données**

#### **A) Manipulations essentielles**

Avant de commencer toute recherche sur notre base de données, il convient de réaliser quelques manipulations qui ont pour objectif de la rendre exploitable.

- Tout d'abord nous renommons les 10 variables avec des noms plus courts qui rendent les manipulations plus faciles, tout en restant explicites pour la lecture.
- Ensuite nous vérifions la nature des variables sous R grâce à la fonction "str".
- Notre base étant déjà "propre" nous n'avons pas besoin de supprimer les pays pour lesquels il nous manque certaines informations (dits "observations incomplètes"), néanmoins il aurait fallu le faire si nous en avions car cela aurait faussé les estimations puisque le nombre de pays aurait différé d'une variable à l'autre.

Nous pouvons désormais commencer l'analyse de notre base de données. Après avoir commenté les statistiques descriptives nous nettoierons notre base de données en deux temps. Dans un premier temps nous identifierons et supprimerons les valeurs atypiques, car leur présence peut biaiser les résultats des estimations économétriques. Dans un second temps nous verrons s'il y a des corrélations entre les variables ; en effet le modèle ne doit pas contenir de variables corrélées entre elles car cela biaiserait également les résultats.



## **B) Statistiques descriptives**

Commençons maintenant à regarder les informations contenues dans chaque variable grâce aux statistiques descriptives obtenues par la fonction “summary”. Cette commande nous donne accès aux valeurs minimales et maximales de chaque variable, la moyenne, la médiane et les quartiles.

D’après l’annexe n°1 nous remarquons d’emblée qu’il existe de fortes disparités entre le pays, l’étendue (écart entre les valeurs minimales et maximales pour une variable donnée) est assez importante, notamment pour les variables PIB par habitant, salaire moyen ou encore taux de chômage. Cet indice étant très sensible aux valeurs extrêmes, nous supposons donc qu’il y aura des valeurs atypiques au moins pour ces 3 variables.

Nous voyons aussi que pour la grande majorité des variables, la moyenne et la médiane sont très proches. La médiane se différencie de la simple moyenne arithmétique car elle suppose que 50% de l’échantillon a des valeurs inférieures à celle-ci, et 50% a des valeurs supérieures. Ainsi, lorsqu’elle est proche de la moyenne - comme c’est le cas ici - cela signifie que les données sont relativement uniformes, il n’y a apparemment pas de valeurs qui faussent la moyenne. Il apparaît donc que nos observations sont relativement homogènes même si l’étendue reste généralement élevée.

Nous pouvons aussi nous arrêter sur la variable “PIB par habitant” qui est la seule ayant une médiane éloignée de la moyenne. 50% des pays ont un PIB par habitant inférieur à 15.613,8\$ alors que la moyenne s’élève à 24.728,2\$, cela signifie qu’un ou plusieurs pays ont un PIB par habitant très fort qui a tendance à “tirer” la moyenne vers le haut. C’est dans ce genre de cas qu’il est intéressant de comparer moyenne et médiane, dont la différence a été expliquée précédemment.

## **C) Valeurs atypiques**

Pour déterminer s’il existe des valeurs atypiques pour chaque variable explicative nous utilisons la fonction “boxplot”. Dans cette représentation graphique les points qui s’affichent au dessus ou en dessous de la moustache correspondent aux individus qui sont

**potentiellement** atypiques. À partir de ces constats, nous réaliserons des tests pour trouver quels pays sont vraiment atypiques afin ensuite de les retirer de la base.

Après observation des boîtes à moustache disponibles en annexe n°2, nous pouvons segmenter les variables en trois groupes :

➤ les variables n'ayant pas de valeurs atypiques : salaire moyen et indice de corruption → il n'y a alors pas de test à effectuer.

➤ les variables ayant potentiellement une seule valeur atypique : PIB par habitant, indice de Gini, IDH, dépenses de santé, et dépense en R et D → il est nécessaire d'appliquer le test de Grubb afin de vérifier que la valeur soit belle et bien atypique.

➤ les variables ayant entre 2 et 10 valeurs atypiques : taux de chômage et espérance de vie à la naissance → il est nécessaire d'appliquer le test de Rosner pour voir quelles sont les valeurs atypiques.

Nous allons faire une analyse complète pour la variable PIB par habitant ainsi que pour le taux de chômage, puis nous résumerons dans un tableau les conclusions obtenues pour chaque autre variable. Les résultats des tests sont disponibles annexes n°3 et 4.

### **1- Test de Grubbs : PIB\_hab**

(1)  $H_0$  : il n'y a pas de variable atypique dans la série

$H_1$  : il y a 1 variable atypique

(2)  $p\text{-value}=0,02695 < \alpha=0,05$

(3) Nous validons donc  $H_1$ , il existe une valeur atypique pour cette variable.

Les fonctions "sort" puis "order" nous révèlent que le PIB par habitant atypique est celui du Luxembourg qui s'élève à 100.428\$. Il est en effet atypique car trop élevé par rapport aux autres pays. Nous décidons donc de retirer l'observation correspondante n°36 de notre base afin que la suite de notre analyse ne soit pas biaisée.

### **2- Test de Rosner : Tx\_chom**

(1)  $H_0$  : il n'y a pas de variable atypique dans la série

$H_1$  : il y a entre 2 et 10 variables atypiques

(2) test statistique > critical statistique pour 3 valeurs

(3) Nous rejetons donc  $H_0$ , il existe 3 valeurs atypiques pour cette variable.

La boîte à moustache nous avait montré 5 valeurs potentiellement atypiques, le test de Rosner quant à lui, nous révèle qu'il y a en réalité 3 valeurs atypiques. Il s'agit des trois pays ayant les taux de chômage les plus élevés : l'Afrique du Sud, la Grèce, et l'Espagne correspondent respectivement aux observations 1,19, et 26 qui sont donc à retirer de la base.

### 3- Tableau récapitulatif des valeurs atypiques

TABLEAU N°3 : Tableau récapitulatif des variables atypiques détectées

	Nombre de valeurs potentiellement atypiques	Test appliqué	Outliers	observations(s) atypique(s) à retirer de la base
Taux de pauvreté	0	X	X	X
PIB par habitant	1	Grubbs	> 100.428	36
Salaire moyen	0	X	X	X
Indice de Gini	1	Grubbs	> 62	1
Indice de corruption	0	X	X	X
IDH	1	Grubbs	< 45	22
Taux de chômage	5	Rosner	> 22,1	19/26/1
Dépenses en santé	1	Grubbs	> 16,7	21
Espérance de vie	2	Rosner	< 62	1
Dépenses R&D	1	Grubbs	> 4,3	30

Le tableau ci-dessus récapitule les résultats trouvés en termes d'étude des valeurs atypiques pour chaque variable. Nous avons donc 9 valeurs atypiques, néanmoins cela ne signifie pas qu'il y a 9 observations à retirer de la base. En effet, étant donné que l'Afrique du

Sud a des valeurs atypiques dans 3 variables différentes, il y aura seulement 7 observations à enlever de la base. Ainsi sur nos 59 observations, 7 sont considérées comme atypiques : il s'agit d'un peu moins de 12% de notre base. Nous créons donc une nouvelle base : "basepauvreté1" ne contenant pas les individus 1, 19, 21, 22, 26, 30, et 36. Ainsi, contrairement aux hypothèses émises dans la partie précédente la variable 'salaire moyen' ne contient pas de valeurs atypiques, en revanche nos intuitions étaient bonnes quant aux variables PIB par habitant et taux de chômage.

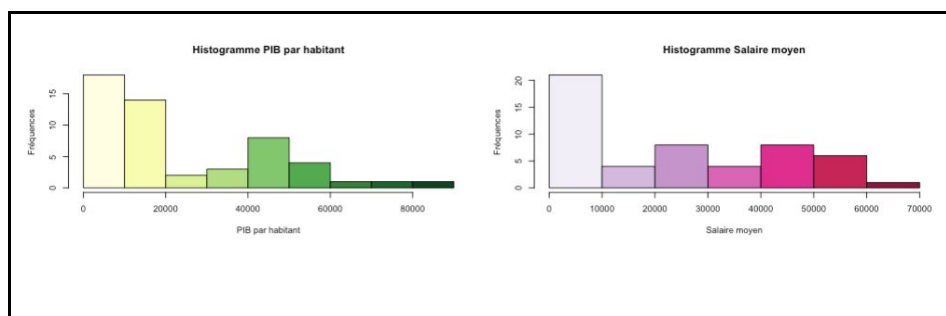
## **D) Corrélation entre les variables explicatives**

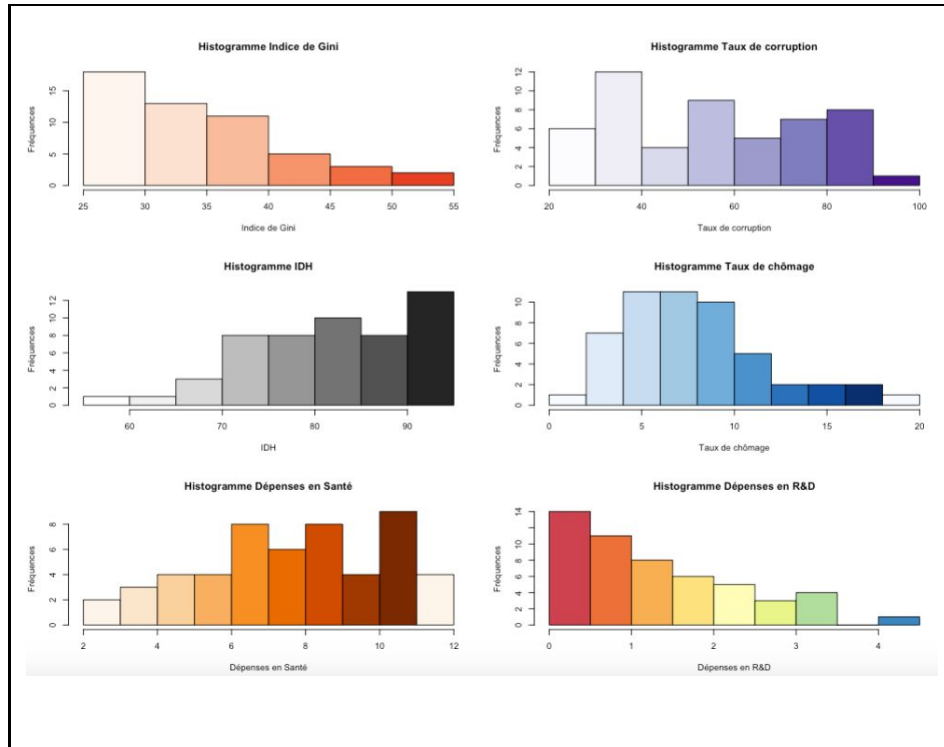
L'une des hypothèses fondamentale de la méthode des Moindres Carrés Ordinaires (MCO) est l'absence de multicolinéarité entre les variables explicatives. En effet si elles sont trop corrélées entre elles, cela fausse le lien qu'elles ont avec la variable à expliquer. Pour vérifier cette corrélation, nous pouvons utiliser le coefficient de Pearson ou de Spearman en fonction de la distribution de chaque variable.

Il convient donc de voir si les variables suivent ou non une loi normale ; si elles suivent toutes une loi normale nous appliquerons le test de Pearson, dans le cas contraire nous prendrons le coefficient de Spearman. Pour ce faire, nous commencerons par regarder visuellement la distribution de chaque variable quantitative **continue** (la variable espérance de vie étant quantitative discrète nous ne l'intégrerons pas dans cette analyse) grâce aux histogrammes, nous pourrons alors émettre des hypothèses que nous confirmerons par la suite par le test de Shapiro.

### **1- Histogrammes des variables explicatives quantitatives continues**

**FIGURE N° 1 : Histogrammes des variables sans les valeurs atypiques**





D'après la figure n°1 et l'annexe n°5 (histogrammes des variables contenant les valeurs atypiques), nous remarquons que les variables ne semblent pas suivre une loi normale. On note cependant des différences de distribution entre les histogrammes provenant de la nouvelle et l'ancienne base. La distribution du taux de chômage s'apparente davantage à la loi normale sur l'histogramme de la nouvelle base, ci-dessus.

Nous pouvons compléter cette analyse en cherchant la **skewness** et la **kurtosis** pour nos variables : ces coefficients donnent une indication sur l'aplatissement et la symétrie de la distribution des variables. La loi normale se caractérise par un coefficient d'aplatissement et d'asymétrie égal à zéro. S'ils sont différents de zéro il y a 2 situations possibles pour chaque coefficient :

- skewness positif : asymétrie à gauche, et skewness négatif : asymétrie à droite
- kurtosis positif : la distribution sera moins aplatie que la dist normale théorique, et s'il est négatif : la distribution sera plus aplatie

Disponibles en annexe n°7 pour la base initiale et en annexe n°6 pour la base sans les valeurs atypiques, les coefficients nous confirment donc l'asymétrie des variables, plus particulièrement pour PIB par habitant, Gini, taux de chômage et IDH qui ont un skewness élevé par rapport à celui des autres variables. Pour les 3 premières variables citées ci-dessus

l'asymétrie est à gauche, en revanche l'asymétrie est à droite pour l'IDH, ce qui confirme nos hypothèses émises au regard des histogrammes. En ce qui concerne le coefficient d'aplatissement, nous constatons que pour toutes les variables leur distribution est moins aplatie que la distribution normale. Ainsi, les variables s'apparentant le plus à une loi normale sont le salaire moyen et le taux de corruption, pour lesquels les deux coefficients sont les plus proches de zéro.

En comparant les histogrammes des variables de l'ancienne et de la nouvelle base nous avons pu constater une différence, spécialement pour le taux de chômage. Nous comparons à présent les 2 coefficients pour cette variable afin de voir si nous notons effectivement une différence écart important. Le coefficient de skewness passe de 1,47 à 0,85 ce qui réduit donc presque de moitié l'asymétrie à gauche. En ce qui concerne la kurtosis qui mesure l'aplatissement de la courbe, l'écart de coefficient entre les 2 bases est aussi notable, cela est largement visible sur les histogrammes de variables. Plus généralement, sans prendre en compte les pays atypiques nous constatons que les 2 coefficients sont plus élevés pour chacune des variables ce qui traduit une plus forte asymétrie à gauche, ainsi qu'une distribution plus concentrée des variables.

## **2- Test de Shapiro**

Pour confirmer ou non ces hypothèses sur la normalité des variables il est nécessaire d'effectuer le test de Shapiro-Wilk. En statistique ce dernier permet de tester l'hypothèse nulle selon laquelle une variable suit une loi normale, c'est à dire que l'échantillon de la variable est issu d'une population normalement distribuée.

Ici encore nous réaliserons l'analyse entière sur la variable PIB par habitant, pour les autres variables vous trouverez le résultat des tests en annexe n°7.

### **→ Shapiro test : PIB\_hab**

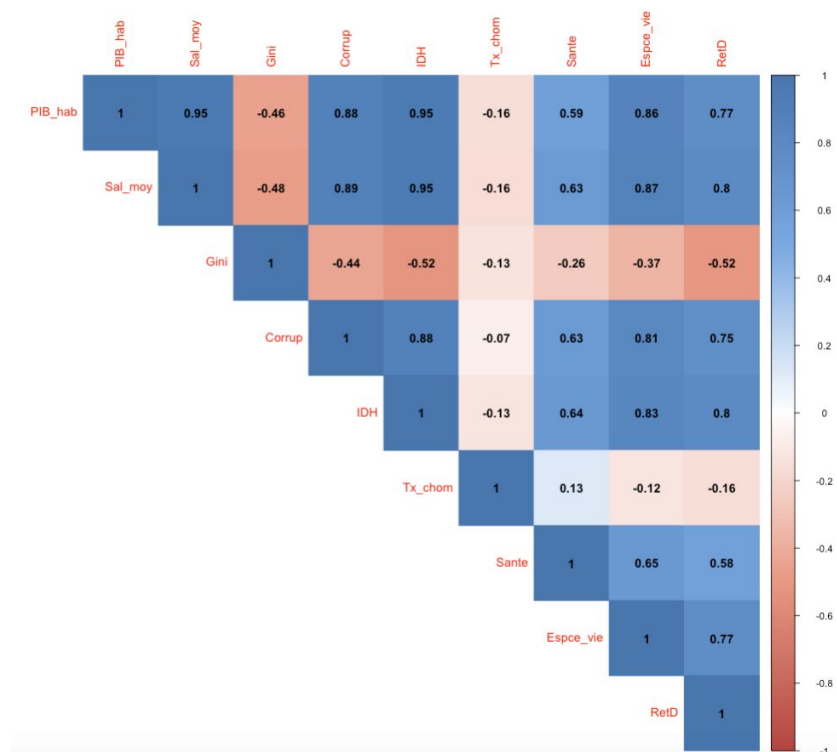
- (1)  $H_0$  : La variable suit une loi normale
- (2)  $p\text{-value}=0,000033 < \alpha=0,05$
- (3) On rejette l'hypothèse  $H_0$ , la variable PIB par habitant ne suit pas une loi normale au seuil de risque de 5%.

D'après l'annexe n°8 nous voyons que seule la variable 'Santé' suit une loi normale, toutes les autres variables ont une p-value inférieure à 0,05 c'est pourquoi nous appliquerons la corrélation de Spearman. La corrélation de Spearman est étudiée lorsque deux variables statistiques semblent corrélées sans que la relation entre les deux variables soit de type affine (ne forme pas une droite).

### 3- Matrice de corrélation de Spearman et ACP

Nous allons à présent analyser la matrice de corrélation en regardant les coefficients, s'ils sont compris entre 0 et 0,5 nous considérons que les variables ne sont pas corrélées, entre 0,5 et 0,6 nous pouvons supposer l'existence d'une corrélation et au delà de 0,5 nous estimons que la corrélation est forte - donc nous supprimerons les variables concernées.

FIGURE N°2 : Matrice de corrélation de Spearman



Au regard de la figure n°2 ou de l'annexe n°9 nous remarquons que la plupart des variables sont fortement corrélées entre elles. Pour les variables PIB par habitant et salaire moyen par exemple la corrélation s'élève à 95%, ce qui est considérable.

Nous voulions comparer cette analyse avec celle de l'analyse en composante principale pour confirmer et/ou compléter les constats faits à partir de la matrice. Disponible en annexe n°10, l'ACP confirme nos constats puisque l'on remarque que la grande majorité des variables sont regroupées tout près de l'axe 1 ; il s'agit du PIB par habitant, du salaire moyen, du taux de corruption, de l'IDH, de l'espérance vie et des dépenses en recherche et développement. Cela confirme donc visuellement ce que nous avons pu voir grâce aux coefficients de corrélation : il existe de fortes corrélations entre les variables énumérées. En revanche pour l'indice de Gini, les dépenses en santé ainsi que le taux de chômage on remarque ici encore qu'elles sont graphiquement excentrées des autres variables, ce qui confirme la corrélation plus faible de celles-ci.

Au vu de ces résultats nous conserverons donc les variables Gini, taux de chômage ainsi que dépenses en Santé qui ont les corrélations les plus faibles. Nous testerons plusieurs modèles jusqu'à ce que l'un d'eux vérifie toutes les hypothèses d'un modèle MCO que nous énumérerons ci-après, pour in fine trouver la meilleure estimation de la pauvreté en 2015.

## **II- Estimations de modèles par MCO**

Les variables étant fortement corrélées entre elles, nous estimerons plusieurs modèles en prenant comme base les 2 ou 3 variables les moins corrélées (indice de Gini, taux de chômage et dépenses en santé), puis testerons des modèles alternatifs en ajoutant des variables un peu plus corrélées mais qui nous paraissent pertinentes afin d'expliquer le taux de pauvreté. Nous cherchons ainsi le meilleur modèle qui réponde à toutes les hypothèses et qui inclut suffisamment de variables qui constitueront pour vous les clefs, les arguments d'une bonne campagne présidentielle. En effet, ce sera sur ces éléments que vous pourrez proposer d'agir pour réduire massivement la pauvreté au Tadjikistan.

De même, nous appliquerons certains modèles sur l'ancienne comme sur la nouvelle base débarrassée des pays atypiques, car tout comme nous avons pu noter des différences significatives pour les coefficients d'asymétrie et d'aplatissement, il nous paraît intéressant de pouvoir noter des différences de validité des hypothèses en prenant en compte ou non les pays atypiques. Nous effectuerons l'analyse complète pour un premier modèle, en



développant chaque test (nature, utilité, application), puis nous résumerons l'ensemble des résultats des tests en paragraphe pour les modèles suivants s'ils utilisent la même méthode d'estimation. Finalement, nous terminerons cette partie par un tableau récapitulatif de tous les modèles estimés, les hypothèses qu'ils valident ou non, et les solutions apportées dans ce cas.

### **A) Hypothèses d'un modèle MCO**

Tout modèle estimé par la méthode des moindres carrés ordinaires se doit de vérifier 5 critères pour pouvoir tirer des conclusions :

➤ les résidus suivent une loi normale : cette condition doit faire l'objet de vérification seulement sur des échantillons de moins de 30 observations, ce n'est pas notre cas donc nous n'aurons pas besoin de vérifier cette première hypothèse

➤ le modèle est bien spécifié : il faut vérifier si la forme fonctionnelle est correcte, c'est à dire si  $E(\varepsilon_i) = 0$

➤ les variables explicatives ne sont pas corrélées entre elles : s'il y avait multicolinéarité cela signifierait que certaines variables mesurent le même phénomène.

➤ les erreurs sont homoscédastiques : la variance du terme d'erreur doit être la même pour chaque observation

➤ les observations ne sont pas influentes : la suppression d'un pays ne doit pas affecter la précision de la régression

Si ces conditions ne sont pas vérifiées alors le modèle n'est pas valide. En fonction de chaque problème nous prendrons des mesures et créerons un nouveau modèle qui puisse répondre à ces 5 critères, et ce, jusqu'à ce que l'on trouve celui qui estimera au mieux le taux de pauvreté en 2015. C'est à partir de ce dernier que l'on tirera nos conclusions qui vous aideront pour votre campagne.

## B) Estimation d'un premier modèle MCO

Nous estimons un premier modèle dans lequel nous supposons que les variables sont toutes exogènes, c'est à dire qu'elles ne dépendent pas du modèle en lui-même, elles ne sont pas explicables par Y, en revanche elles tentent de l'expliquer. Nous retenons ici seulement les 3 variables les moins corrélées :

$$Tx\_pauv_i = \alpha + \beta_1 * Gini_i + \beta_2 * Tx\_chom_i + \beta_3 * Sante_i + u_i$$

### MODÈLE N°1 : Indice de Gini, taux de chômage et dépenses en santé par MCO

```
> summary(modele)

Call:
lm(formula = Tx_pauv ~ Gini + Tx_chom + Sante, data = basepauvreté1)

Residuals:
    Min       1Q   Median       3Q      Max
-14.9540  -3.4898  -0.5732   2.3185  13.7090

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -15.44689    5.43408  -2.843  0.00655 **
Gini          0.69265    0.11682   5.929 3.22e-07 ***
Tx_chom       0.88545    0.20608   4.297 8.41e-05 ***
Sante        -0.08787    0.33394  -0.263  0.79358
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.726 on 48 degrees of freedom
Multiple R-squared:  0.5176,    Adjusted R-squared:  0.4875
F-statistic: 17.17 on 3 and 48 DF,  p-value: 1.037e-07
```

Le modèle empirique est donc le suivant :

$$\widehat{Tx\_pauv}_i = -15,45 + 0,69 * \widehat{Gini}_i + 0,89 * \widehat{Tx\_chom}_i - 0,09 * \widehat{Sante}_i$$

### 1- La forme fonctionnelle du modèle est-elle correcte ?

Le premier critère consiste à vérifier si le modèle est sous la bonne forme fonctionnelle. Pour cela nous utilisons le test de Ramsey. L'idée de ce test est de voir si le modèle souffre de l'omission d'une ou plusieurs variables pertinentes en introduisant une variable fictive. Il consiste donc à vérifier la significativité de l'effet de la variable fictive introduite. Si elle n'est pas significative, alors la spécification du modèle est complète, c'est-à-dire que le modèle a pris en compte toutes les variables pertinentes qui interviennent

dans l'explication de la variable à expliquer. En revanche, si la variable fictive est significative, alors des variables susceptibles d'influencer les variations de Y devront être introduites.

Nous posons les hypothèses suivantes :

(1)  $H_0$  : la forme fonctionnelle du modèle est linéaire

$H_1$  : la forme fonctionnelle du modèle n'est pas linéaire

(2) Résultat du test n°1: test RESET

```
> reset(modele)

RESET test

data: modele
RESET = 0.10962, df1 = 2, df2 = 46, p-value = 0.8964
```

(3) Nous voyons ci-dessus que la p-value associée au test de Ramsey est supérieure à 5% (p-value = 0,8964), notre modèle a donc une forme fonctionnelle linéaire au seuil de risque de 5%.

## 2- Existe-t-il de la multicollinéarité entre les variables ?

Le second critère consiste à vérifier l'existence d'un problème de multicollinéarité entre les variables explicatives. Pour cela nous utilisons le critère du "Variance Inflation Factor" (VIF). La règle de décision est la suivante : si le VIF est supérieur à 10 alors il y a un problème de multicollinéarité.

Résultat du test n°2: test VIF

```
> vif(modele)
      Gini Tx_chom  Sante
1.043180 1.029769 1.063050
```

Le VIF de chacune des variables est inférieur à 10, il n'y a donc aucun problème de multicollinéarité, nous pouvons alors passer au test suivant.

## 3- Les résidus sont-ils homoscedastiques ?

Nous pouvons désormais étudier une troisième caractéristique. D'après les hypothèses fondamentales des MCO, les résidus du modèle doivent être homoscedastiques, c'est-à-dire que la variance du terme d'erreur doit être la même pour chaque observation. Pour vérifier si

cette hypothèse est vérifiée nous utilisons le test de Breusch-Pagan qui vérifie si toutes les variables explicatives, excepté la constante, ont un effet significatif sur la variable à expliquer. Dans ce cas, la variance de l'erreur est constante pour toutes les estimations. Nous posons les hypothèses suivantes :

(1)  $H_0$  : les résidus du modèle sont homoscedastiques

$H_1$  : les résidus du modèle sont hétéroscédastiques

(2) Résultat du test n°3: test de Breush-Pagan

```
> bptest(modele)

studentized Breusch-Pagan test

data: modele
BP = 8.9696, df = 3, p-value = 0.0297
```

(3) La p-value étant supérieure à 0,01 nous pourrions accepter  $H_0$  au seuil de risque de 1%, cependant nous préférons considérer que les résidus sont hétéroscédastiques pour que l'interprétation de ce modèle, si nous le gardons comme modèle final, soit la plus exacte possible. Nous allons donc nous intéresser aux candidats potentiels de cette hétéroscédasticité.

→ Il y a 4 causes possibles de l'hétéroscédasticité :

- 1re hypothèse : la forme fonctionnelle ne convient pas
- 2e hypothèse : cela est lié aux valeurs atypiques
- 3e hypothèse : cela est lié aux variables explicatives
- 4e hypothèse : la cause n'est pas identifiable

Ce ne sont pas les 2 premières hypothèses puisque nous les avons déjà vérifiées, en revanche il est probable que le problème d'hétéroscédasticité soit liée au fait que les résidus et les variables explicatives sont corrélés, nous procéderons donc à une "recherche de coupable" en 2 étapes. Premièrement nous réaliserons des graphiques qui nous donneront une idée des coupables potentiels, puis nous confirmerons cela en réalisant un test statistique sur nos variables explicatives quantitatives.

#### 4- Y a-t-il des coupables de l'hétéroscédasticité du modèle?

Résultat de test n°4 : Test recherche de coupable, "residualPlots"

```
> residualPlots(modele)
      Test stat Pr(>|Test stat|)
Gini      -0.5672      0.5733
Tx_chom    0.8271      0.4124
Sante     -0.0506      0.9598
Tukey test -0.0830      0.9339
```

D'après l'annexe n°12, nous pouvons supposer que l'indice de Gini et le taux de chômage sont responsables de l'hétéroscédasticité des résidus car les courbures sont légèrement arrondies et non parfaitement horizontales. Il faut cependant vérifier si elles sont significatives par un test statistique donné par la commande "residualPlots". L'hypothèse nulle selon laquelle les courbures ne sont pas significatives (variables non "coupables") est acceptée pour l'ensemble des variables de ce modèle.

Ainsi, aucune variable explicative du modèle ne peut expliquer l'hétéroscédasticité des résidus. Généralement lorsque les coupables ne sont pas connus, nous utilisons la méthode de White. Néanmoins le problème de cette méthode est qu'elle nécessite un nombre important d'observations ( $\geq 250$ ), il faudra donc être vigilant sur les résultats car ce n'est pas le cas ici.

TABLEAU N°4 : Résultats de la méthode de White

```
> coeftest(modele, vcov=vcovHC(modele,type="HC3"))

t test of coefficients:

      Estimate Std. Error t value Pr(>|t|)
(Intercept) -15.446891   5.603386 -2.7567 0.0082313 **
Gini          0.692650   0.093874  7.3785 1.938e-09 ***
Tx_chom       0.885446   0.218495  4.0525 0.0001846 ***
Sante        -0.087869   0.412943 -0.2128 0.8323945
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```

> waldtest(modele, vcov=vcovHC(modele,type="HC3"))
Wald test

Model 1: Tx_pauv ~ Gini + Tx_chom + Sante
Model 2: Tx_pauv ~ 1
  Res.Df Df      F    Pr(>F)
1      48
2      51 -3 33.995 6.226e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

TABLEAU N°5 : Comparaison des écarts-types issus des MCO avec ceux de White

	Écarts-types MCO	Écarts-types White
<b>Indice de Gini</b>	0,11	0,09
<b>Taux de chômage</b>	0,2	0,21
<b>Dépenses de santé</b>	0,33	0,41

Nous constatons d'emblée que le modèle estimé par la méthode de White est significatif au seuil de risque de 5%, comme de 1%. N'ayant pas de qualité d'ajustement générale mesurée par  $R^2$ , nous décidons de comparer ses écarts-types avec ceux de la méthode des MCO, le but étant de minimiser les écarts-types. Nous remarquons que l'écart-type de l'indice de Gini diminue de 2 points de pourcentage lorsque l'on estime par White. À l'inverse, les écarts types augmentent pour les variables taux de chômage et dépenses de santé, plus particulièrement pour la variable dépenses de santé. Si l'on se concentre sur ce critère, il est donc plus intéressant de retenir la méthode des MCO que celle de White.

Finalement, ce modèle analysé par la méthode des MCO souffre de plusieurs problèmes : celui d'hétéroscédasticité des résidus, le fait que la variable 'dépenses de santé' ne soit pas significative et que le  $R^2$  ne soit pas très élevé. Nous allons donc formuler un nouveau modèle en y intégrant une nouvelle variable explicative pour voir si cela améliore la qualité de l'analyse de la variable expliquée, tout en respectant les conditions de MCO.

## 5- Même modèle estimé à partir de l'ancienne base ("modelebis")

⇒ Avant cela regardons rapidement les résultats des tests pour ce même modèle en prenant la base de départ contenant les 7 pays atypiques. D'après l'annexe n°13, nous remarquons d'emblée que la qualité d'ajustement estimée par le  $R^2$  est plus faible dans ce nouveau modèle, ce qui prouve que les variables atypiques ajoutées n'améliorent pas la qualité de prédiction du modèle, voire elles réduisent la précision des estimations des coefficients et des valeurs prédites. Malgré cela nous décidons de continuer la vérification des hypothèses car la partie qui nous intéresse est celle concernant l'hétéroscédasticité des résidus et ses potentiels coupables. La forme fonctionnelle est validée au seuil de 5% (p-value=0,07494) mais nous ne pouvons pas savoir s'il s'agit de la meilleure forme fonctionnelle étant donné que la valeur est très proche de 0,05. Encore une fois nous décidons de continuer l'analyse de ce modèle. Concernant la multicolinéarité des variables, les coefficients du VIF sont aussi très faibles ( $<2$ ), nous validons donc l'absence de multicolinéarité entre les variables. Enfin, nous vérifions l'homoscédasticité des erreurs grâce au test de Breush-Pagan (BP) : la p-value étant inférieure à 0,05 nous rejetons cette hypothèse et cherchons les potentiels coupables.

TABLEAU N°6 : Résultats du test de recherche de coupables

	Test stat	Pr(> Test stat )
Gini	-2.0733	0.04293 *
Tx_chom	-1.8174	0.07471 .
Sante	0.0274	0.97821
Tukey test	-2.3217	0.02025 *

D'après le tableau n°5 nous voyons qu'il s'agit dans ce cas de l'indice de Gini pour lequel la curvature est significative au seuil de risque de 5%. Il est à noter que la 'variable taux de chômage' est aussi potentiellement responsable car sa p-value de 0,075 est très proche de 0,05. Au vu de ces résultats nous allons tester 2 modèles alternatifs, l'un en retirant la variable 'indice de Gini' et le second en retirant la variable 'taux de chômage' afin de déterminer le rôle qu'ont ces 2 variables dans l'explication de l'hétéroscédasticité.

$$\rightarrow Tx\_pauv_i = \alpha + \beta_1 * Tx\_chom_i + \beta_2 * Sante_i + u_i$$

```
> bptest(modeletestgini)

studentized Breusch-Pagan test

data: modeletestgini
BP = 2.2567, df = 2, p-value = 0.3236
```

$$\rightarrow Tx\_pauv_i = \alpha + \beta_1 * Gini_i + \beta_2 * Sante_i + u_i$$

```
> bptest(modeletestchom)

studentized Breusch-Pagan test

data: modeletestchom
BP = 5.1203, df = 2, p-value = 0.07729
```

Ces deux tests ont pour but de vérifier s'il y a toujours de l'hétéroscédasticité lorsque l'on enlève les variables qui semblent avoir un impact sur ce problème. Nous tenons à préciser que les deux modèles vérifient les conditions de la méthode MCO mais que leurs  $R^2$  sont bien trop faibles ( $R^2 < 0,2$  pour les deux modèles) pour qu'ils soient considérés comme pertinents pour représenter la variation du taux de pauvreté.

Nous remarquons d'emblée que le modèle duquel nous avons retiré la variable indice de Gini vérifie l'hypothèse d'homoscédasticité des résidus, c'est également le cas pour celui duquel nous avons retiré la variable 'taux de chômage' même si c'est plus juste car la p-value est égale à 0,078. Cela semble logique puisque dès le départ cette variable n'était coupable qu'au seuil de 10%, donc lorsque l'on estime un modèle sans celle-ci mais comprenant l'indice de Gini qui lui, semble agir beaucoup sur le problème d'hétéroscédasticité, le modèle passe le test de BP de justesse. Il apparaît donc que la variable 'indice de Gini' est celle qui contribue le plus à l'hétéroscédasticité des résidus lorsque l'on prend la base contenant les pays atypiques. Nous supposons donc que l'hétéroscédasticité c'est à dire, la différence de variance entre les variables, peut être expliquée en majorité par l'Afrique du Sud, observation que nous avons retirée dans la nouvelle base car considérée comme atypique pour taux de chômage, espérance de vie et indice de Gini. Sachant cela nous allons formuler un deuxième modèle ne comprenant pas cette dernière variable, tout en ajoutant l'indice de corruption.



### C) Estimation d'un deuxième modèle MCO ("modele2")

Nous estimons un deuxième modèle où nous supposons ici encore que les variables sont toutes exogènes. Le modèle empirique de cette 2ème estimation est le suivant :

$$Tx\_pauv_i = \alpha + \beta_1 * Tx\_chom_i + \beta_2 * Sante_i + \beta_3 * Corrup_i + u_i$$

#### MODÈLE N°2 : Taux de chômage, dépenses en santé et taux de corruption par MCO

```
> summary(modele2)

Call:
lm(formula = Tx_pauv ~ Tx_chom + Sante + Corrup, data = basepauvreté1)

Residuals:
    Min       1Q   Median       3Q      Max
-16.1633  -2.7686   0.4771   4.2108  11.4438

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  18.35458    3.39760   5.402 2.02e-06 ***
Tx_chom       0.40891    0.23697   1.726  0.0909 .
Sante        1.15428    0.48400   2.385  0.0211 *
Corrup      -0.28119    0.05779  -4.866 1.27e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.168 on 48 degrees of freedom
Multiple R-squared:  0.4404,    Adjusted R-squared:  0.4054
F-statistic: 12.59 on 3 and 48 DF,  p-value: 3.399e-06
```

Le modèle empirique est le suivant :

$$\widehat{Tx\_pauv}_i = 18,35 + 0,41 * \widehat{Tx\_chom}_i + 1,15 * \widehat{Sante}_i - 0,28 * \widehat{Corrup}_i$$

Nous observons tout d'abord que dans ce modèle de régression la p-value associée au test de Fisher qui nous informe de la significativité du modèle, est inférieure à 0,05, ce qui traduit le fait qu'il est pertinent de l'analyser. Nous pouvons également voir que la qualité d'ajustement, mesurée par le R<sup>2</sup> ajusté, a baissé par rapport à notre première estimation MCO passant de 0,49 à 0,41. Nous pouvons à présent regarder si le modèle vérifie les conditions de validité d'un MCO. Disponibles en annexe n°14, les résultats des tests nous montrent que le modèle a une forme fonctionnelle linéaire au seuil de risque de 5% (p-value=0,81) et que les variables ne sont pas corrélées entre elles (VIF<2). Néanmoins, malgré le fait que nous ayons retiré l'indice de Gini qui semblait expliquer en partie le problème d'hétéroscédasticité, la

p-value du test de BP reste inférieure à 0,05 (p-value=0,001). De plus, nous notons que cette valeur est encore plus faible que celle du modèle n°1 où elle était de 0,02.

Nous procédons donc à la recherche de coupable en regardant les courbures et leur significativité au seuil de risque de 5% (cf. annexe n°14). Cependant, nous nous retrouvons dans la même situation que pour le modèle n°1 où nous n'avions pas trouvé de coupables. Nous estimons encore une fois ce modèle par la méthode de White qui s'applique lorsque l'on ne trouve pas de coupables potentiels.

TABLEAU N°7 : Estimation du modèle n°2 par la méthode de White

```

> coeftest(modele2, vcov=vcovHC(modele2,type="HC3"))

t test of coefficients:

              Estimate Std. Error t value Pr(>|t|)
(Intercept) 18.354580   4.562698  4.0227 0.0002029 ***
Tx_chom      0.408908   0.199695  2.0477 0.0460863 *
Sante        1.154277   0.490482  2.3534 0.0227488 *
Corrup       -0.281194   0.056913 -4.9407 9.861e-06 ***
---
> waldtest(modele2, vcov=vcovHC(modele2,type="HC3"))
Wald test

Model 1: Tx_pauv ~ Tx_chom + Sante + Corrup
Model 2: Tx_pauv ~ 1
      Res.Df Df    F    Pr(>F)
1         48
2         51 -3 14.344 8.415e-07 ***
---

```

D'après le tableau n°6 ci-dessus, nous constatons que le test de White ne réduit l'écart type que de la variable taux de chômage pour les deux autres, l'écart type est quasiment le même si l'on compare à la méthode des MCO.

Finalement, ce modèle ne permet pas de résoudre les problèmes que nous avons déjà avec le premier modèle. De plus, du fait de la forte corrélation qu'ont toutes les variables explicatives entre elles (exceptés le taux de chômage et les dépenses de santé), nous sommes obligés de les garder et ne pouvons en ajouter qu'une seule autre, afin que notre modèle valide l'hypothèse de non-colinéarité entre les variables.

### **III- Estimations de modèles par DMC**

#### **A) Estimation d'un premier modèle DMC ("modele3")**

Les modèles 1 et 2 n'étant pas concluants du fait de l'hétéroscédasticité, nous décidons de créer un troisième modèle dans lequel nous introduisons le salaire moyen dans notre régression. En effet, cette variable est particulièrement intéressante car si le modèle vérifie toutes les conditions, que le salaire moyen et le modèle dans sa globalité sont significatifs, alors vous saurez que le niveau du salaire moyen constitue un déterminant essentiel du taux de pauvreté. De plus, c'est une variable sur laquelle il est relativement facile d'agir si nous la comparons avec le PIB par habitant ou le taux de corruption par exemple. Il sera donc plus aisé pour vous de prendre des mesures allant dans ce sens pour réduire la pauvreté de votre pays.

Dans la partie économique nous avons supposé que le salaire moyen est une variable endogène. Or si l'on applique la méthode des MCO les paramètres estimés seront biaisés et inconsistants du fait de cette endogénéité. Nous pouvons cependant obtenir des paramètres consistants en utilisant la méthode des DMC qui est valable dans le cas où il y a plus de variables instrumentales que de variable(s) endogène(s), comme dans celui où le nombre est le même.

#### **MODÈLE N°3 : Salaire moyen, taux de chômage et dépenses en santé par DMC**

```

> summary(reg_DMC,vcov=sandwich, diagnostics = TRUE)

Call:
ivreg(formula = Tx_pauv ~ Sal_moy + Tx_chom + Sante | Tx_chom +
      Sante + RetD + Gini, data = basepauvret  1)

Residuals:
    Min       1Q   Median       3Q      Max
-17.4325  -3.5674   0.6824   4.4661  10.5949

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  8.395e+00  4.094e+00   2.051  0.04579 *
Sal_moy      -5.010e-04  8.749e-05  -5.726 6.55e-07 ***
Tx_chom      -3.096e-02  2.436e-01  -0.127  0.89941
Sante        2.376e+00  7.035e-01   3.377  0.00146 **

Diagnostic tests:
              df1 df2 statistic  p-value
Weak instruments  2  47   30.862 2.76e-09 ***
Wu-Hausman       1  47   10.193 0.00252 **
Sargan           1 NA     1.365 0.24263
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.597 on 48 degrees of freedom
Multiple R-Squared:  0.3599,    Adjusted R-squared:  0.3199
Wald test: 20.56 on 3 and 48 DF, p-value: 1.041e-08

```

Dans cette premi  re estimation par DMC nous avons choisi d'utiliser 2 variables instrumentales pour le salaire moyen : les d  penses en R&D ainsi que l'indice de Gini car nous pensons que ces variables sont corr  l  es    notre variable endog  ne et nous supposons qu'elles sont ind  pendantes du terme d'erreur du taux de pauvret  . Dans ce type de r  gression, avant toute interpr  tation, il faut au pr  alable v  rifier les r  sultats de trois tests, si l'un des trois n'est pas v  rifi  , nous ne pourr  ons exploiter les r  sultats de l'estimation par DMC. Nous avons donc 3 tests    v  rifier dans le cadre de cette premi  re estimation par DMC.

### 1- Les instruments sont-ils pertinents ?

Premi  rement nous regardons la pertinence des instruments choisis par le test de Weak Instruments, il nous informe en effet de la corr  lation entre les instruments et la variable endog  ne. Aussi nous posons les hypoth  ses suivantes :

(1)  $H_0$  : les instruments ne sont pas significatifs

$H_1$  : les instruments sont significatifs car suffisamment corr  l  s avec la variable endog  ne

(2) La p-value de ce test  $\approx 0 < 0,05$  donc nous acceptons l'hypoth  se 1 selon laquelle les instruments utilis  s sont pertinents.

## 2- La variable est-elle endogène ?

Regardons maintenant si le salaire moyen est effectivement endogène ou pas, cette réponse nous sera donnée par le test de Wu-Hausman. Les hypothèses sont les suivantes :

(1)  $H_0$  : la variable testée est exogène

$H_1$  : la variable testée est endogène

(2) La p-value associée à ce test =  $0,0025 < 0,05$  : nous rejetons  $H_0$  au seuil de risque de 5% : le salaire moyen ayant pour instrument IDH et dépenses en R&D est bel et bien exogène. Nous pouvons donc interpréter librement les résultats de cette estimation.

## 3- Les instruments sont-ils valides ?

Finalement, le troisième et dernier test à interpréter pour un modèle DMC est celui de Sargan qui indique si les instruments utilisés sont corrélés avec les termes d'erreurs de l'équation, c'est à dire s'ils sont valides. Ce test doit être appliqué seulement lorsqu'il y a plus d'instruments que de variables supposées endogènes comme c'est le cas ici. Nous posons :

(1)  $H_0$  : les instruments sont valides

$H_1$  : les instruments ne sont pas valides

(2) La p-value étant  $0,24 > 0,05$  nous acceptons  $H_0$ , les instruments sont bien valides au seuil de risque de 5%.

Les instruments de ce premier modèle sont donc valides et pertinents, et le salaire moyen est effectivement endogène lorsque l'on utilise 2 instruments. En revanche, la qualité d'ajustement du modèle étant relativement faible (à peine plus de 30%) nous préférons nous intéresser à d'autres variables, afin de voir si elles expliquent mieux le taux de pauvreté.

## **B) Estimation d'un deuxième modèle DMC ("modele4")**

Après avoir étudié un modèle contenant le taux de chômage, la santé ainsi que le salaire moyen où cette dernière variable nous dévoilait un aspect microéconomique du taux de pauvreté, nous décidons de passer à une approche plus macroéconomique en substituant l'IDH au salaire moyen. L'IDH, s'il est significatif, donne en réalité 3 clefs d'interventions

concrètes dans l'économie puisqu'il englobe le PIB par habitant, l'espérance de vie à la naissance ainsi que la durée moyenne de scolarisation. Ainsi, nous espérons répondre le mieux possible à vos attentes concernant la réduction la pauvreté.

#### MODÈLE N°4 : IDH, taux de chômage et dépenses en santé par DMC

```
> summary(reg_DMC2,vcov=sandwich, diagnostics = TRUE)

Call:
ivreg(formula = Tx_pauv ~ IDH + Tx_chom + Sante | Tx_chom + Sante +
Corrup, data = basepauvreté1)

Residuals:
    Min       1Q   Median       3Q      Max
-14.694  -3.701   0.629   3.512  14.380

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  73.6900    13.3226   5.531 1.29e-06 ***
IDH          -0.9094     0.1877  -4.846 1.36e-05 ***
Tx_chom       0.3157     0.1956   1.614 0.11315
Sante         1.7280     0.5524   3.128 0.00299 **

Diagnostic tests:
              df1 df2 statistic p-value
Weak instruments  1 48   63.080 2.69e-10 ***
Wu-Hausman       1 47    2.515  0.119
Sargan           0 NA        NA      NA
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.123 on 48 degrees of freedom
Multiple R-Squared: 0.4485,    Adjusted R-squared: 0.414
Wald test: 17.27 on 3 and 48 DF, p-value: 9.645e-08
```

Nous allons procéder à la même analyse que pour le modèle précédent estimé aussi par DMC, c'est à dire en regardant les valeurs des tests Weak Instruments (WI) et Hausman. Nous n'aurons toutefois pas besoin de vérifier la validité des instruments par le test de Sargan puisqu'ici nous avons sélectionné un instrument (taux de corruption) pour une variable supposée endogène (IDH). La p-value très proche de 0 du test WI nous informe de la pertinence de l'instrument utilisé, celle de Hausman quant à elle remet en cause l'endogénéité supposée de l'IDH. En effet nous acceptons l'hypothèse nulle selon laquelle l'IDH est en fait exogène : nous pouvons donc estimer ce modèle par la méthode des MCO puis, comme pour tout modèle MCO, nous vérifierons les conditions de validité.

#### **1- Estimation du même modèle par MCO ("modele4bis")**

Le modèle est le suivant :  $Tx\_pauv_i = \alpha + \beta_1 * IDH_i + \beta_2 * Tx\_chom_i + \beta_3 * Sante_i + u_i$

#### MODÈLE "4bis" : IDH, taux de chômage et dépenses en santé par MCO

```

> summary(modele4bis)

Call:
lm(formula = Tx_pauv ~ IDH + Tx_chom + Sante, data = basepauvreté1)

Residuals:
    Min       1Q   Median       3Q      Max
-13.5047  -3.6305   0.8387   3.2730  14.8184

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  59.0503     9.2824   6.362 7.01e-08 ***
IDH          -0.6942     0.1290  -5.380 2.19e-06 ***
Tx_chom       0.4344     0.2254   1.927  0.0599 .
Sante         1.2116     0.4608   2.630  0.0115 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.953 on 48 degrees of freedom
Multiple R-squared:  0.4787,    Adjusted R-squared:  0.4461
F-statistic: 14.69 on 3 and 48 DF,  p-value: 6.444e-07

```

Disponibles en annexe n°15, les résultats des tests de validité des conditions d'un modèle MCO nous indiquent plusieurs choses. Tout d'abord la forme fonctionnelle est valide puisque la valeur associée à ce test est de 0.518, nous notons aussi les VIF très faibles (<2) qui signalent l'absence de multicolinéarité entre les variables. De plus, nous voyons que les erreurs sont homoscedastiques car la p-value du test de Breush Pagan s'élève à 0,3 ; nous pouvons enfin vérifier la dernière condition d'un modèle MCO : l'influence des données mesurée par la distance de Cook. En effet, cette dernière mesure l'effet de la suppression d'une donnée, ainsi les points ayant une distance de Cook supérieure à 1 seront considérés comme méritant un examen plus approfondi dans l'analyse. Au regard du graphique des distances de Cook disponible en annexe n°15 avec les autres tests, nous constatons qu'aucunes données ne semblent influentes car elles sont comprises entre 0 et 0,08. Ce modèle répond donc à toutes les conditions d'un bon modèle estimé par la méthode des MCO. Cependant nous remarquons que le taux de chômage ne semble pas avoir un impact significatif sur le taux de pauvreté. Nous décidons donc de passer cette variable en logarithme en espérant que cela augmente sa significativité pour pouvoir la considérer dans les interprétations ci-après.



## 2- Estimation du même modèle par MCO avec taux de chômage en logarithme (“modele4ter”)

Le modèle est le suivant :  $Tx\_pauv_i = \alpha + \beta_1 * IDH_i + \beta_2 * \log(Tx\_chom)_i + \beta_3 * Sante_i + u_i$

MODÈLE “4ter” : IDH, taux de chômage en logarithme et dépenses en santé par MCO

```
> summary(modele4ter)

Call:
lm(formula = Tx_pauv ~ IDH + log(Tx_chom) + Sante, data = basepauvretél)

Residuals:
    Min       1Q   Median       3Q      Max
-14.0476  -3.4550   0.7834   3.3313  14.4039

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  58.5111     9.2499   6.326 7.96e-08 ***
IDH          -0.7169     0.1248  -5.744 6.16e-07 ***
log(Tx_chom)  3.1010     1.5078   2.057  0.0452 *
Sante         1.1936     0.4585   2.603  0.0123 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.924 on 48 degrees of freedom
Multiple R-squared:  0.4838,    Adjusted R-squared:  0.4516
F-statistic: 15 on 3 and 48 DF, p-value: 5.101e-07
```

Nous pouvons voir que le fait d’avoir passé le taux de chômage en logarithme l’a rendu significatif au seuil de 5%. De plus ce modèle passe tous les tests (cf. résultats des tests en annexe n°16), il répond aux 5 conditions d’un modèle MCO et a une qualité d’ajustement relativement bonne par rapport aux modèles précédents.

## IV- Tableaux récapitulatifs des modèles créés et interprétation des résultats

TABLEAU N°8 : Tableau récapitulatif des modèles MCO

	Modèle n°1	Modèle n°2	Modèle “4bis”	Modèle “4ter”
Variables estimées	Gini Tx_chom Sante	Corrup Tx_chom Sante	IDH Tx_chom Sante	IDH log(Tx_chom) Sante
Ramsey	✓	✓	✓	✓
VIF	✓	✓	✓	✓



<b>Breush-Pagan</b>	✗	✗	✓	✓
<b>Coupables</b>	0	0		
<b>Cook</b>			✓	✓
<b>R<sup>2</sup></b>	0,52	0,44	0,48	0,48
<b>R<sup>2</sup> ajusté</b>	0,49	0,41	0,45	0,45
<b>Nombre de variables significatives *</b>	2/3	2/3	2/3	3/3

\* : nombre de variables significatives au seuil de 5%.

TABLEAU N°9 : Tableau récapitulatif des modèles DMC

	<b>Modèle n°3</b>	<b>Modèle n°4</b>
<b>Variables estimées</b>	Sal_moy (endogène) Tx_chom Sante	IDH (endogène) Tx_chom Sante
<b>Nombre d'instrument(s)</b>	2	1
<b>Weak instruments</b>	✓	✓
<b>Hausman *</b>	✓	✗
<b>Sargan</b>	✓	
<b>R<sup>2</sup></b>	0,36	0,42
<b>R<sup>2</sup> ajusté</b>	0,32	0,45
<b>Nombre de variables significatives</b>	2/3	2/3

\* : ok si bonne supposition c'est à dire si la variable supposée endogène l'est vraiment

Au vu de ces résultats nous voyons que le meilleur modèle que nous avons pu obtenir est le modèle "4ter" : celui-ci est issu d'une régression initialement effectuée par la méthode des DMC puisque l'on soupçonnait l'IDH d'être endogène. Finalement le test de Hausman nous a montré que cette variable est en fait exogène, c'est pourquoi nous avons par la suite estimé ce même modèle par la méthode des MCO. Sachant qu'il respecte toutes les conditions de validité d'un MCO et que son coefficient de détermination (R<sup>2</sup>) est l'un des plus fort que nous

avons pu trouver, nous avons décidé de le garder comme modèle final après avoir amélioré la significativité de la variable ‘taux de chômage’ en l’ayant passé en logarithme. De plus, comme dit précédemment dans la partie où nous estimions ce modèle pour la première fois, l’IDH contient 3 dimensions, donc il offre trois fois plus de possibilités en termes de mesures politiques et économiques. Nous allons donc passer à l’interprétation de ce modèle “4ter”.

➤ Rappelons que le modèle théorique est le suivant :

$$Tx\_pauv_i = \alpha + \beta_1 * IDH_i + \beta_2 * \log(Tx\_chom)_i + \beta_3 * Sante_i + u_i$$

➤ Et le modèle empirique est le suivant :

$$\widehat{Tx\_pauv}_i = 58,51 - 0,72 * \widehat{IDH}_i + 3,1 * \log(\widehat{Tx\_chom}_i) + 1,2 * \widehat{Sante}_i$$

Nous observons premièrement que dans ce modèle de régression à trois variables explicatives, la p-value est inférieure à 0,05 : le modèle est donc significatif, c’est à dire qu’il est pertinent de l’analyser.<sup>27</sup>

Ensuite nous pouvons également voir que le  $R^2$  est égal à 0,48 ce qui n’est pas très élevé. Celui-ci se définit comme la part de variation de la variable  $Y_i$  qui est expliquée par les variations des variables explicatives (souvent exprimées en %). Aussi, plus il est proche de 1 plus le modèle est de bonne qualité, c’est à dire que les variables explicatives prises en compte expliquent à elles seules une grande partie de la variation de Y.

Avec ce modèle, 48% des variations du taux de pauvreté sont expliquées, ce qui n’est pas très satisfaisant.

Nous notons également que l’estimateur de la variable dépenses de santé est positif (1,19) : ainsi, lorsque les dépenses de santé augmentent de 1% le taux de pauvreté augmentera de 1,19% au seuil de risque de 5%. Cette relation est inverse à celle que l’on avait trouvée avec le premier modèle MCO. L’estimateur de la variable était négatif, néanmoins sa p-value était supérieure à 0,05 donc la variable était non interprétable. Ici, tout comme dans le second modèle, la p-value est inférieure à 0,05 et l’estimateur est négatif. Cela nous laisse finalement penser que plus les dépenses de santé dans un pays sont élevées, plus il a de chance d’avoir un fort taux de pauvreté, ce qui contredit notre intuition de départ. En effet de

<sup>27</sup> Voir page 48 : MODÈLE “4ter” : IDH, taux de chômage en logarithme et dépenses en santé par MCO

nombreuses études et articles ont montré que la bonne santé d'une population est un facteur de sortie de la pauvreté, et qu'à l'inverse, une population en mauvaise santé a plus de chance d'être pauvre, on peut donc en conclure que la variable dépense de santé n'est pas représentative de l'état de santé réel d'une population.

La p-value de la variable IDH est extrêmement faible ce qui traduit l'impact fort qu'à cette variable sur l'explication du niveau de pauvreté. Cela est sans doute dû à la nature de la variable qui se mesure à partir du produit intérieur brut (PIB) par habitant, l'espérance de vie des citoyens d'un Etat et le niveau d'éducation. En intégrant trois différents facteurs dans sa grille de lecture, cet indice permet de mesurer sous plusieurs angles le développement des Etats et de leur population. Ainsi nous avons pu voir que l'Indice de Développement Humain est un facteur particulièrement significatif quant à la lutte contre la pauvreté.

Il convient à présent d'analyser le coefficient du taux de chômage, en faisant attention car nous l'avons passé en logarithme, cela change donc la manière de l'interpréter. Le coefficient estimé de cette variable étant égal à 3,1, cela signifie que lorsque le taux de croissance de l'IDH est égal à 1, alors le taux de pauvreté devrait augmenter de 3,1%. Il apparaît donc que l'impact d'une modification sur le taux de pauvreté est important, cependant la significativité de cette variable est assez limitée car la valeur du test de student (qui informe de la significativité d'une variable) est de 0,0452 ce qui est tout juste inférieur à  $\alpha = 0,05$ .

## **PARTIE 3 : Conclusion et discussion des résultats**

Le but de cette étude était de répondre aux besoins de la future présidente du Tadjikistan pour savoir quels sont les principaux déterminants du taux de pauvreté et quelles peuvent être les clefs pour en entreprendre une éventuelle éradication.

Nous avons donc répondu à cela en cherchant un modèle contenant les variables les plus faciles d'accès : c'est à dire celles qui seraient les plus faciles à modifier une fois aux commandes de ce pays. Certaines contraintes économétriques telles que l'absence de corrélation entre les variables explicatives, nous a obligé à réduire le nombre de variables en retenant 2 variables principales (taux de chômage et dépenses en santé), auxquelles nous avons ajouté d'autres variables qui nous semblaient pertinentes vis à vis des possibilités que vous aurez si vous êtes élue. Nous avons donc testé plusieurs modèles et avons retenu celui qui ajoutait l'indice de développement humain aux 2 variables citées précédemment.

Nous pouvons à présent vous proposer quelques exemples de mesures pouvant être mises en place pour chacune des composantes du taux de pauvreté retenues dans notre modèle final. Nous ne donnerons pas de mesures pour les dépenses en santé : en effet, nous avons vu que le coefficient estimé de ce paramètre est positif. Aussi, augmenter les dépenses en santé du gouvernement ne réduit apparemment pas la pauvreté, au contraire, d'après notre modèle, elle augmente avec les dépenses. À l'opposé, nous ne proposons pas non plus de mesures pour réduire ces dépenses, car cela nous paraît être un non-sens économique puisque l'état de santé d'une population est tout de même fondamental. En revanche, nous pouvons soumettre quelques clefs de réduction de la pauvreté en regardant les variables 'IDH' et 'taux de chômage'.

Commençons par nous intéresser au niveau de développement humain d'un pays. La première composante de celui-ci est le PIB par habitant : nous pouvons associer les mesures possibles avec celles du taux de chômage car elles ont toutes deux pour but d'augmenter l'activité économique générale du pays par le biais du retour au plein-emploi. Le problème du taux de chômage se trouve dans la dualité du marché de l'emploi : il existe une opposition entre les "*insiders*" (travailleurs qualifiés, rémunérés correctement, hauts coûts de

licenciement) et les “*outsiders*” (chômeurs, travailleurs peu qualifiés, conditions de travail précaires, salaires plus faibles etc.). Ainsi, des mesures allant dans le sens de réduction du taux de chômage pour augmenter l’activité économique et le PIB par habitant seraient : aider à l’employabilité des chômeurs en mettant en place des primes de première embauche dans les entreprises, en proposant des formations, ou encore en facilitant le *matching* entre les employeurs et les employés en homogénéisant les coûts de licenciement entre les *insiders* et les *outsiders*.

Ces propositions concernaient la première composante de l’IDH : le PIB par habitant ainsi que le taux de chômage, à présent nous pouvons nous pencher sur la deuxième composante de l’IDH qu’est l’espérance de vie. Cette dernière étant corrélée avec les dépenses en Santé pour lesquelles nous avons trouvé un effet contraire à l’effet initialement attendu, nous ne proposerons pas de mesures correspondantes afin de rester cohérent avec notre modèle final.

En revanche pour la dernière composante, c’est à dire, l’éducation mesurée à travers de la durée moyenne de scolarisation, nous pouvons vous présenter certains moyens d’améliorer l’Éducation au sein de votre pays. Vous pourriez par exemple ouvrir de nouvelles écoles en périphérie, là où l’éducation est plus rare (car mode de vie plus rural et position géographique désavantageuse puisque loin des écoles), proposer davantage de bourses au mérite en éducation secondaire, ou encore entreprendre des campagnes de sensibilisation chez les jeunes pour accroître leurs connaissances et leur esprit critique. Si de telles mesures portent leurs fruits, alors elles amélioreront à terme l’activité économique du pays en réduisant le taux de chômage du fait de la plus grande qualification des travailleurs, amplifiant ainsi le PIB moyen par habitant puisque nous attendons que les salaires suivent cette hausse de qualification.

Nous sommes conscients du budget important qu’impliquent ces mesures, cependant, si elles corrigent effectivement le taux de pauvreté, votre pays entrera dans un cercle vertueux dans lequel l’activité économique permettra à terme, de rembourser ces apports financiers. Il faut donc le voir comme un investissement sur le long terme.

Enfin, nous pouvons terminer cette conclusion en nous interrogeant sur la chose suivante : étant donné que le modèle que nous avons retenu explique seulement la moitié de

la variation du taux de pauvreté, nous sommes en droit de nous demander quelles variables influencent l'autre moitié ?

Grâce aux graphiques de corrélation étudiés dans la partie économique, nous avons pu observer trois variables supplémentaires qui semblent efficaces dans la lutte contre la pauvreté. Tout d'abord les variables liées au niveau de revenu des individus telles que le PIB par habitant et le salaire moyen, semblent être très fortement et négativement corrélées au taux de pauvreté.<sup>28</sup> Ainsi, employer des mesures qui permettent d'accroître le salaire moyen des individus ou de manière générale leur richesse, les aideront efficacement à sortir de la pauvreté. De plus, la relation entre la pauvreté et la corruption semble être parfaitement linéaire et croissante <sup>29</sup>, combattre la corruption est donc un bon moyen de lutte contre la pauvreté dans un pays.

---

<sup>28</sup> Voir graphique n°1 et n°2 page 11 et 12

<sup>29</sup> Voir graphique n°4 page 14

# **PARTIE 4 : Table des matières**

<b>PARTIE 1 : Analyse économique</b>	<b>2</b>
I- Abstract	2
II- Introduction	3
III- Analyse économique des variables retenues	6
Une variable à expliquer	7
Et 9 variables explicatives	8
1- Pib par habitant ( $X_1$ ) et salaire moyen ( $X_2$ )	10
2- Indice de Gini ( $X_3$ )	13
3- Indice de corruption ( $X_4$ )	14
4- Indice de développement humain ( $X_5$ )	15
5- Taux de chômage ( $X_6$ )	16
6- Dépenses de Santé ( $X_7$ ) et espérance de vie à la naissance ( $X_8$ )	17
7- Dépenses intérieures brutes de recherche et développement ( $X_9$ )	20
<b>PARTIE 2 : Analyse économétrique</b>	<b>23</b>
I- Nettoyage de la base de données	23
Manipulations essentielles	23
Statistiques descriptives	24
Valeurs atypiques	24
1- Test de Grubbs : PIB_hab	25
2- Test de Rosner : Tx_chom	25
3- Tableau récapitulatif des valeurs atypiques	26
D) Corrélation entre les variables explicatives	27
1- Histogrammes des variables explicatives quantitatives continues	27
2- Test de Shapiro	29
II- Estimations de modèles par MCO	31
Hypothèses d'un modèle MCO	32
Estimation d'un premier modèle MCO	32
1- La forme fonctionnelle du modèle est-elle correcte ?	33
2- Existe-t-il de la multicollinéarité entre les variables ?	34
3- Les résidus sont-ils homoscedastiques ?	34
4- Y a-t-il des coupables de l'hétéroscédasticité du modèle?	35
5- Même modèle estimé à partir de l'ancienne base ("modelebis")	37
C) Estimation d'un deuxième modèle MCO ("modele2")	39
III- Estimations de modèles par DMC	41
Estimation d'un premier modèle DMC ("modele3")	41
1- Les instruments sont-ils pertinents ?	42
2- La variable est-elle endogène ?	43

3- Les instruments sont-ils valides ?	43
B) Estimation d'un deuxième modèle DMC ("modele4")	44
1- Estimation du même modèle par MCO ("modele4bis")	45
2- Estimation du même modèle par MCO avec taux de chômage en logarithme ("modele4ter")	46
IV- Tableaux récapitulatifs des modèles créés et interprétation des résultats	46
<b>PARTIE 3 : Conclusion et discussion des résultats</b>	<b>50</b>
<b>PARTIE 4 : Table des matières</b>	<b>53</b>
<b>PARTIE 5 : Annexes</b>	<b>55</b>
Annexe n°1 : Statistiques descriptives de la base de données	55
Annexe n°2 : Diagrammes en boîte des variables	55
Annexe n°3 : Résultat du test de Grubbs pour la variable PIB_hab	55
Annexe n°4 : Résultat du test de Rosner pour la variable Tx_chom	56
Annexe n°5 : Histogrammes des variables issues de la base avec les valeurs atypiques	57
Annexe n°6 : Coefficients d'asymétrie et d'aplatissement sans les valeurs atypiques	57
Annexe n°7 : Coefficients d'asymétrie et d'aplatissement avec les valeurs atypiques	57
Annexe n°8 : Résultats du test de Shapiro appliqué sans les valeurs atypiques	57
Annexe n°9 : Résultats du test de Shapiro appliqué avec les valeurs atypiques	58
Annexe n°10 : Représentation graphique des liens entre les variables avec la méthode de Spearman	58
Annexe n°11 : Analyse des composantes principales de nos variables explicatives	59
Annexe n°12 : Curvatures associées au modèle n°1 MCO	59
Annexe n°13 : Estimation et résultats des tests du premier modèle MCO appliqué sur la base avec les valeurs atypiques	60
Annexe n°14 : Résultats des tests de l'estimation d'un 2ème modèle MCO	61
Annexe n°15 : Résultats des tests MCO à partir du modèle n°4	62
Annexe n°16 : Résultats des tests pour MCO à partir du modèle n°4 avec log(tx_chom)	62
<b>PARTIE 6 : Bibliographie</b>	<b>63</b>



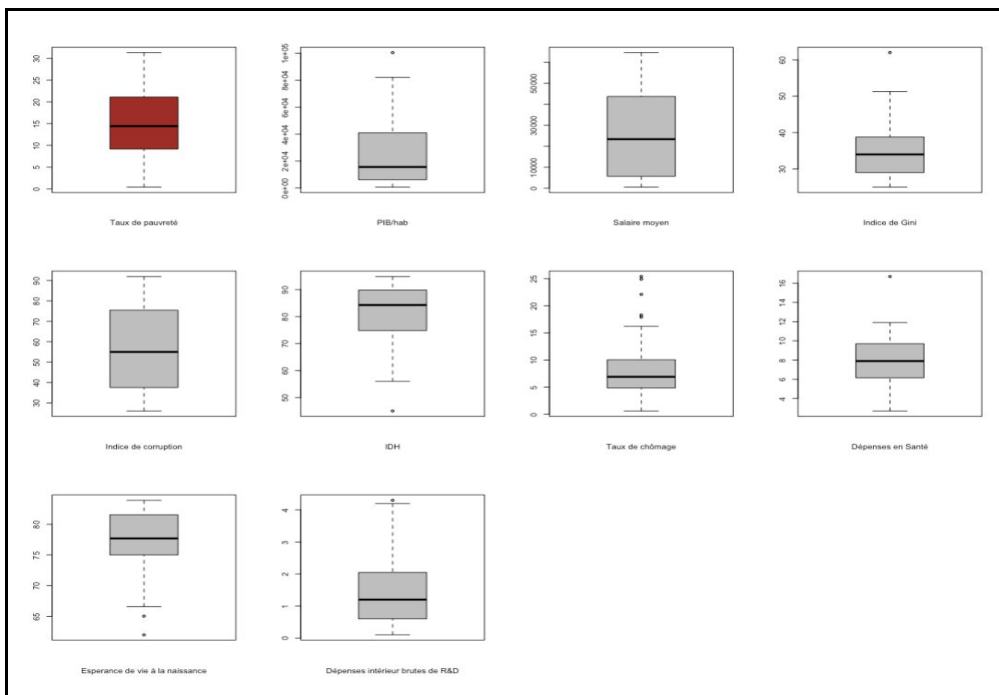
## PARTIE 5 : Annexes

### Annexe n°1 : Statistiques descriptives de la base de données

```
> summary(basepauvreté)
...1          Continents      Tx_pauv      PIB_hab      Sal_moy      Gini
Length:59    Length:59      Min.   : 0.40    Min.   : 639.3    Min.   : 478.7    Min.   :25.00
Class :character  Class :character  1st Qu.: 9.15    1st Qu.: 6201.7    1st Qu.: 5643.2    1st Qu.:29.00
Mode  :character  Mode  :character  Median :14.40    Median : 15613.8    Median :23344.0    Median :34.00
Mean   :15.12    Mean   : 24728.2    Mean   :24873.5    Mean   :34.98
3rd Qu.:21.10    3rd Qu.: 40917.5    3rd Qu.:43674.5    3rd Qu.:38.80
Max.   :31.30    Max.   :100428.0    Max.   :64551.5    Max.   :62.00

Corrup      IDH      Tx_chom      Sante      Espce_vie      RetD
Min.   :26.00    Min.   :45.00    Min.   : 0.600    Min.   : 2.700    Min.   :62.00    Min.   :0.100
1st Qu.:37.50    1st Qu.:74.85    1st Qu.: 4.850    1st Qu.: 6.150    1st Qu.:75.00    1st Qu.:0.600
Median :55.00    Median :84.30    Median : 6.900    Median : 7.900    Median :77.70    Median :1.200
Mean   :56.41    Mean   :82.03    Mean   : 8.468    Mean   : 7.829    Mean   :77.48    Mean   :1.403
3rd Qu.:75.50    3rd Qu.:89.85    3rd Qu.:10.050    3rd Qu.: 9.700    3rd Qu.:81.55    3rd Qu.:2.050
Max.   :92.00    Max.   :94.90    Max.   :25.400    Max.   :16.700    Max.   :83.90    Max.   :4.300
```

### Annexe n°2 : Diagrammes en boîte des variables



### Annexe n°3 : Résultat du test de Grubbs pour la variable PIB\_hab

```
> grubbs.test(`basepauvreté`$"PIB_hab",type=10, two.sided = TRUE)

Grubbs test for one outlier

data: basepauvreté$PIB_hab
G = 3.33760, U = 0.80462, p-value = 0.02695
alternative hypothesis: highest value 100428 is an outlier
```

```
> sort(`basepauvreté`$"PIB_hab")
[1] 639.3044 929.0959 1356.6677 2124.6627 3331.6951 3598.9709 3607.0000 3756.3838
[9] 3859.8144 5406.7039 5585.1178 5733.0000 5840.0469 6124.4916 6175.8760 6227.5867
[17] 8033.3880 8814.0000 9313.7875 9799.3675 10510.7719 10948.0000 11299.0000 11780.1312
[25] 12572.0000 13574.0000 13639.0000 13789.0000 14291.0000 15613.7643 16182.0000 17412.0000
[33] 17715.0000 18167.0000 19252.0000 20887.0000 25817.0000 27105.0000 30170.0000 34524.0000
[41] 35855.0000 36613.0000 40441.0000 40441.0520 41394.0000 42494.0000 43495.0000 44176.0000
[49] 44472.0000 45175.0000 50832.0000 52428.0000 53254.0000 56748.0000 56803.0000 61808.0000
[57] 74521.0000 82081.0000 100428.0000
> order(`basepauvreté`$"PIB_hab")
[1] 22 54 39 58 27 17 4 25 56 40 50 1 55 18 12 42 11 8 49 37 33 57 14 15 43 10 34 3 35 59 45 20
[33] 46 26 44 51 19 13 31 32 30 24 7 47 2 23 9 6 48 41 52 29 16 5 21 28 38 53 36
```

## Annexe n°4 : Résultat du test de Rosner pour la variable Tx\_chom

```
> rosnerTest(`basepauvreté`$Tx_chom, k = 5, alpha = 0.05)

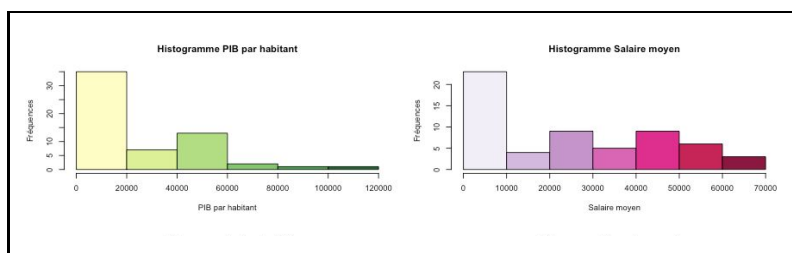
Results of Outlier Test
-----

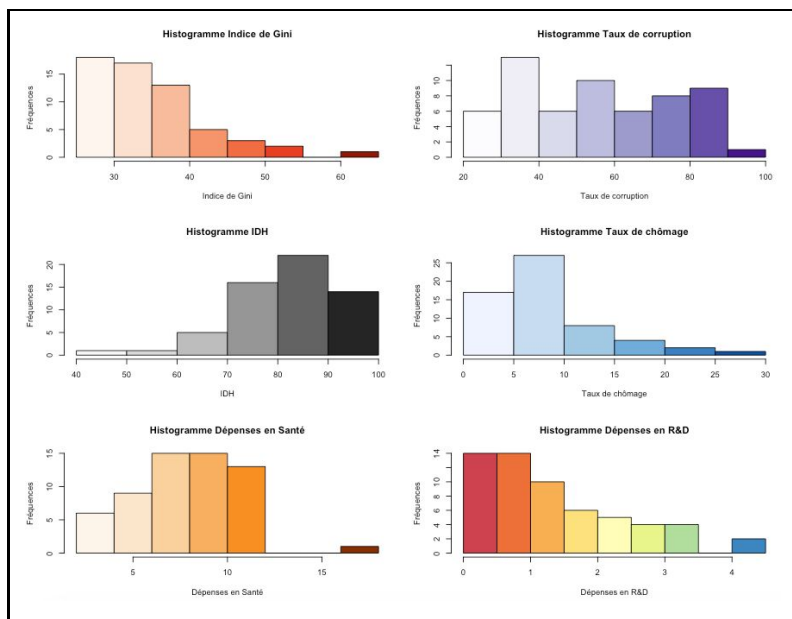
Test Method: Rosner's Test for Outliers
Hypothesized Distribution: Normal
Data: basepauvreté$Tx_chom
Sample Size: 59
Test Statistics: R.1 = 3.200389
R.2 = 3.459825
R.3 = 3.287721
R.4 = 2.728213
R.5 = 2.855111
Test Statistic Parameter: k = 5
Alternative Hypothesis: Up to 5 observations are not
from the same Distribution.
Type I Error: 5%
Number of Outliers Detected: 3

i Mean.i SD.i Value Obs.Num R.i+1 lambda.i+1 Outlier
1 0 8.467797 5.290670 25.4 1 3.200389 3.193214 TRUE
2 1 8.175862 4.833810 24.9 26 3.459825 3.186628 TRUE
3 2 7.882456 4.324437 22.1 19 3.287721 3.179900 TRUE
4 3 7.628571 3.911508 18.3 4 2.728213 3.173022 FALSE
5 4 7.434545 3.665516 17.9 50 2.855111 3.165989 FALSE

> sort(`basepauvreté`$Tx_chom)
[1] 0.6 2.0 3.0 3.1 3.4 3.6 3.6 3.6 4.0 4.3 4.5 4.6 4.6 4.6 4.8 4.9 5.0 5.2 5.3
[20] 5.3 5.6 5.7 6.1 6.2 6.2 6.2 6.7 6.8 6.9 6.9 7.4 7.5 7.5 7.6 8.3 8.4 8.5 9.0
[39] 9.1 9.1 9.4 9.7 9.9 9.9 10.2 10.4 11.4 11.5 11.9 12.4 13.1 14.1 15.2 16.2 17.9 18.3 22.1
[58] 24.9 25.4
> order(`basepauvreté`$Tx_chom)
[1] 55 22 42 37 32 13 18 39 29 38 27 2 11 40 53 33 46 30 21 48 49 6 5 10 16 20 36 47 9 41 52 43
[33] 59 3 12 8 7 51 35 58 23 14 28 34 57 24 54 45 31 44 17 25 56 15 50 4 19 26 1
```

## Annexe n°5 : Histogrammes des variables issues de la base avec les valeurs atypiques





#### Annexe n°6 : Coefficients d'asymétrie et d'aplatissement **sans** les valeurs atypiques

```
> round(sapply(basepauvreté[,4:12],skewness),2)
PIB_hab  Sal_moy    Gini  Corrup    IDH  Tx_chom  Sante Espce_vie  RetD
  0.95    0.38    0.65   0.15  -0.63   0.85   -0.28   -0.50   0.84
> round(sapply(basepauvreté[,4:12],kurtosis),2)
PIB_hab  Sal_moy    Gini  Corrup    IDH  Tx_chom  Sante Espce_vie  RetD
  2.99    1.66    2.74   1.66   2.89   3.32   2.17    2.44   2.80
```

#### Annexe n°7 : Coefficients d'asymétrie et d'aplatissement **avec** les valeurs atypiques

```
> round(sapply(basepauvreté[,4:12],skewness),2)
PIB_hab  Sal_moy    Gini  Corrup    IDH  Tx_chom  Sante Espce_vie  RetD
  1.19    0.38    1.12   0.16  -1.19   1.47   0.31   -1.00   0.88
> round(sapply(basepauvreté[,4:12],kurtosis),2)
PIB_hab  Sal_moy    Gini  Corrup    IDH  Tx_chom  Sante Espce_vie  RetD
  3.98    1.72    4.60   1.67   4.81   5.08   3.62    3.78   2.94
```

#### Annexe n°8 : Résultats du test de Shapiro appliqué **sans** les valeurs atypiques

```
data: basepauvreté$PIB_hab      data: basepauvreté$Sal_moy
W = 0.86673, p-value = 3.311e-05  W = 0.87637, p-value = 6.435e-05

data: basepauvreté$Gini        data: basepauvreté$Corrup
W = 0.94103, p-value = 0.01233    W = 0.9261, p-value = 0.003177

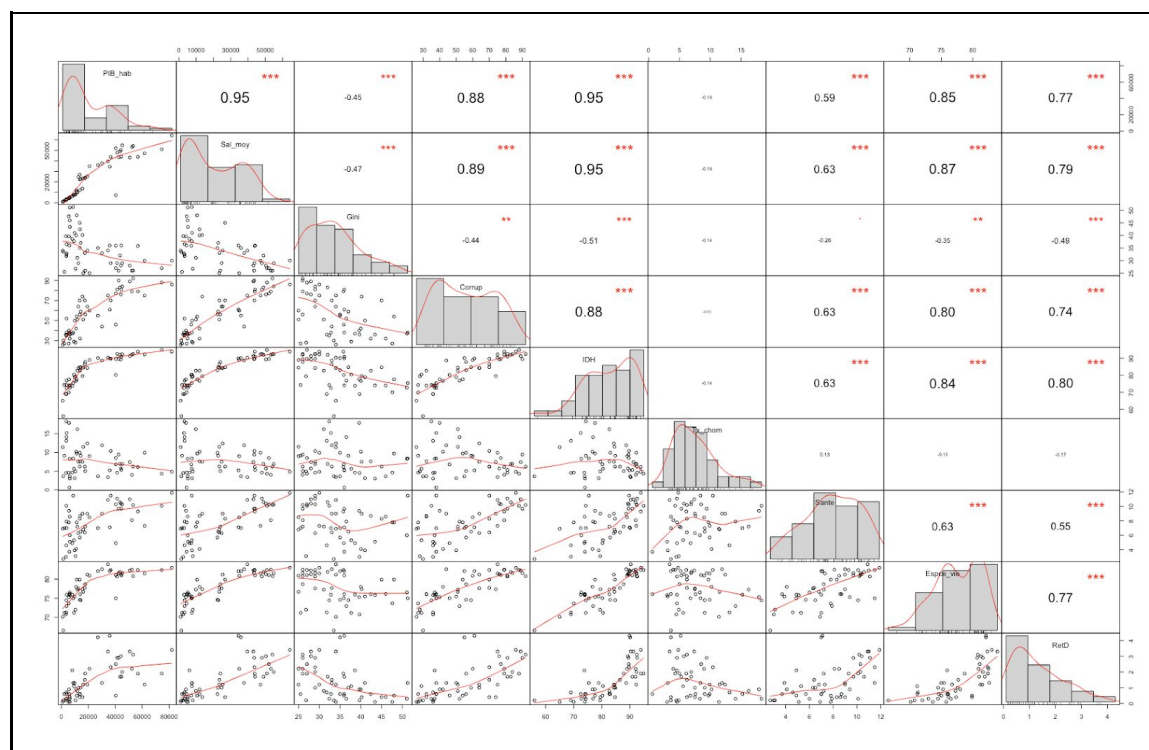
data: basepauvreté$IDH         data: basepauvreté$Tx_chom
W = 0.94139, p-value = 0.01276    W = 0.94273, p-value = 0.01447

data: basepauvreté$Sante       data: basepauvreté$RetD
W = 0.96656, p-value = 0.1505     W = 0.91283, p-value = 0.001028
```

## Annexe n°9 : Résultats du test de Shapiro appliqué **avec** les valeurs atypiques

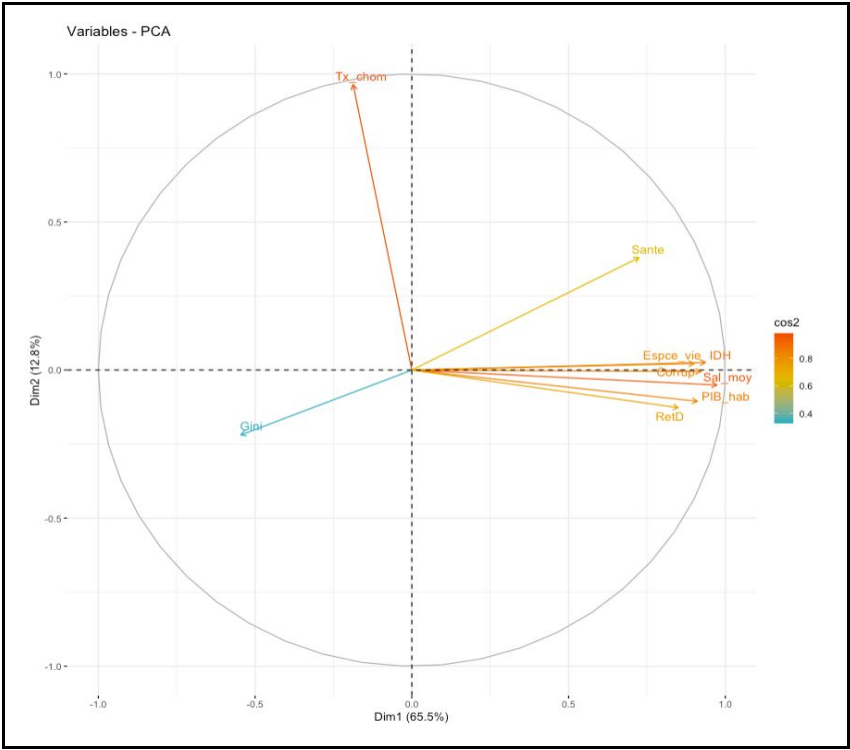
data: basepauvreté\$PIB_hab W = 0.86473, p-value = 9.894e-06	data: basepauvreté\$Sal_moy W = 0.89089, p-value = 7.109e-05
data: basepauvreté\$Gini W = 0.92116, p-value = 0.000952	data: basepauvreté\$Corrup W = 0.9321, p-value = 0.002693
data: basepauvreté\$IDH W = 0.89987, p-value = 0.0001477	data: basepauvreté\$Tx_chom W = 0.86608, p-value = 1.09e-05
data: basepauvreté\$Sante W = 0.97257, p-value = 0.203	data: basepauvreté\$RetD W = 0.90749, p-value = 0.0002812

## Annexe n°10 : Représentation graphique des liens entre les variables avec la méthode de Spearman

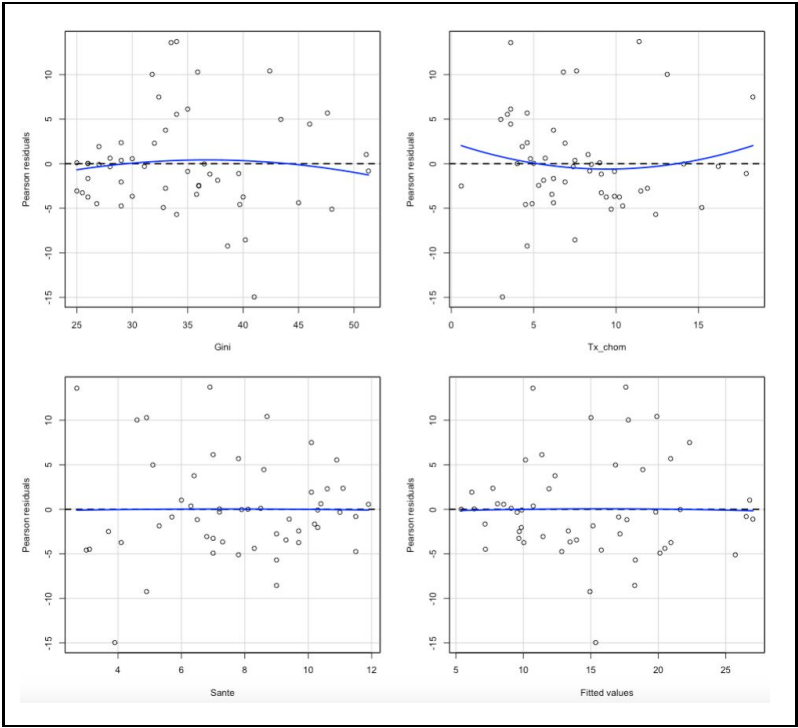




**Annexe n°11 : Analyse des composantes principales de nos variables explicatives**



**Annexe n°12 : Curvatures associées au modèle n°1 MCO**



## Annexe n°13 : Estimation et résultats des tests du premier modèle MCO appliqué sur la base avec les valeurs atypiques

```
> summary(modelebis)

Call:
lm(formula = Tx_pauv ~ Gini + Tx_chom + Sante, data = basepauvreté)

Residuals:
    Min       1Q   Median       3Q      Max
-16.700  -3.389  -1.109   3.462  15.550

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -6.7752     4.7613  -1.423   0.1604
Gini           0.5674     0.1096   5.177 3.28e-06 ***
Tx_chom       0.3637     0.1566   2.323  0.0239 *
Sante        -0.1325     0.3087  -0.429  0.6695
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.176 on 55 degrees of freedom
Multiple R-squared:  0.403,    Adjusted R-squared:  0.3705
F-statistic: 12.38 on 3 and 55 DF,  p-value: 2.691e-06
```

```
> reset(modelebis)

RESET test

data: modelebis
RESET = 2.722, df1 = 2, df2 = 53, p-value = 0.07494
```

```
> vif(modelebis)
      Gini Tx_chom  Sante
1.034488 1.043967 1.035293
```

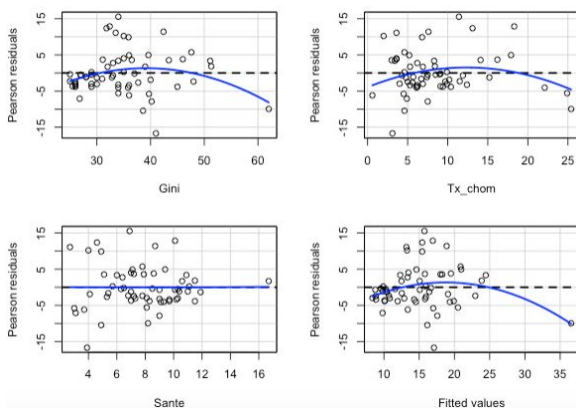
```
> bptest(modelebis)

studentized Breusch-Pagan test

data: modelebis
BP = 9.1896, df = 3, p-value = 0.02687
```

```
> residualPlots(modelebis)

Test stat Pr(>|Test stat|)
Gini      -2.0733      0.04293 *
Tx_chom    -1.8174      0.07471 .
Sante       0.0274      0.97821
Tukey test -2.3217      0.02025 *
```



## Annexe n°14 : Résultats des tests de l'estimation d'un 2ème modèle MCO

```
> reset(modele2)
```

RESET test

data: modele2

RESET = 0.21647, df1 = 2, df2 = 46, p-value = 0.8062

```
> vif(modele2)
```

	Tx_chom	Sante	Corrup
VIF	1.173736	1.924945	1.928583

```
> bptest(modele2)
```

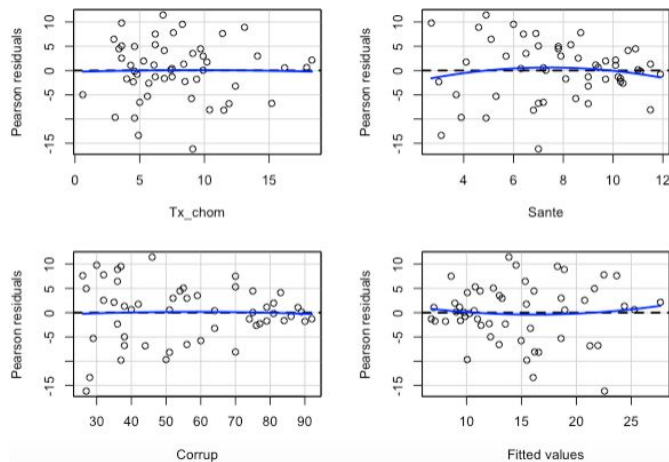
studentized Breusch-Pagan test

data: modele2

BP = 15.535, df = 3, p-value = 0.001412

```
> residualPlots(modele2)
```

	Test stat	Pr(> Test stat )
Tx_chom	-0.0889	0.9295
Sante	-0.7729	0.4434
Corrup	-0.1641	0.8703
Tukey test	0.5277	0.5977



```
> coeftest(modele2, vcov=vcovHC(modele2,type="HC3"))
```

t test of coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	18.354580	4.562698	4.0227	0.0002029	***
Tx_chom	0.408908	0.199695	2.0477	0.0460863	*
Sante	1.154277	0.490482	2.3534	0.0227488	*
Corrup	-0.281194	0.056913	-4.9407	9.861e-06	***

```
> waldtest(modele2, vcov=vcovHC(modele2,type="HC3"))
```

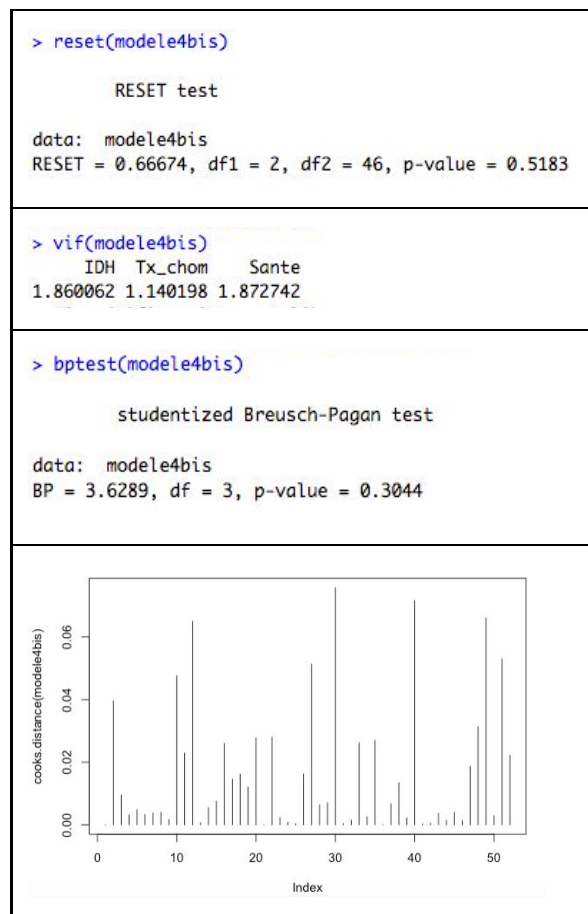
Wald test

Model 1: Tx\_pauv ~ Tx\_chom + Sante + Corrup

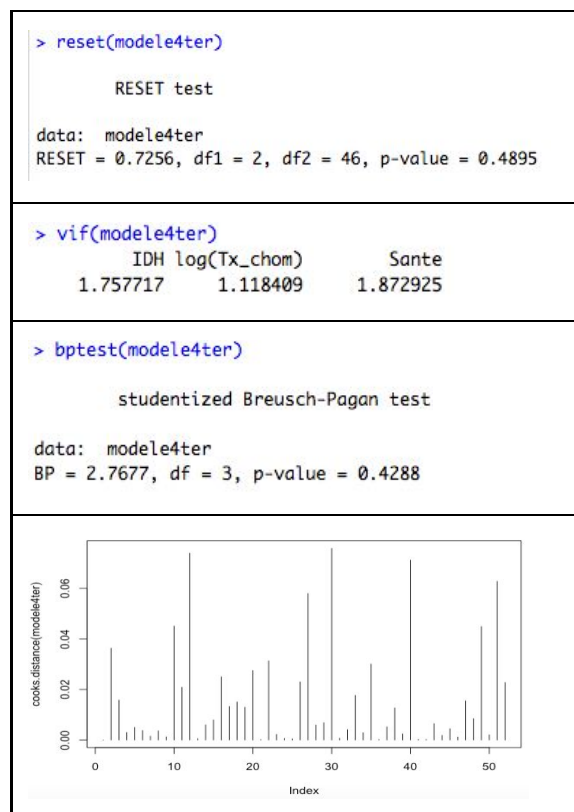
Model 2: Tx\_pauv ~ 1

	Res.Df	Df	F	Pr(>F)
1	48			
2	51	-3	14.344	8.415e-07 ***

## Annexe n°15 : Résultats des tests MCO à partir du modèle n°4



## Annexe n°16 : Résultats des tests pour MCO à partir du modèle n°4 avec log(tx\_chom)





## **PARTIE 6 : Bibliographie**

<sup>1,2</sup> CARTO “Les effets de la crise sur la société : la Pauvreté dans le Monde”, *Le Monde en Carte*, juin 2012, numéro 11, p12-22.

<http://www.sciencespo.fr/ceri/fr/carto/dossier-carto-la-pauvrete-dans-le-monde> (consulté le 24/09/2019)

<sup>3</sup> Alvaredo.F, Chancel.L, Piketty.T, Saez.E, Zucman.G, “Le Rapport sur les inégalités mondiales 2018”, décembre 2017

<http://ses.ens-lyon.fr/actualites/rapports-etudes-et-4-pages/rapport-sur-les-inegalites-mondiales-2018-wid-world-decembre-2017> (consulté le 26/09/2019)

<sup>4,5</sup> Site de l'INSEE, définition du taux de pauvreté, consultable en ligne

<https://www.insee.fr/fr/metadonnees/definition/c1320> (consulté le 29/09/2019)

<sup>6,7</sup> Site de Wikipédia, “Seuil de pauvreté relative” *Seuil de pauvreté*

[https://fr.wikipedia.org/wiki/Seuil\\_de\\_pauvret%C3%A9](https://fr.wikipedia.org/wiki/Seuil_de_pauvret%C3%A9) (consulté le 30/09/2019)

<sup>8</sup> Förster.M et Pearson.M, “Distribution des revenus et pauvreté dans la zone de l'OCDE : tendances et déterminants”, *Revue économique de l'OCDE*, 2002 (numéro 34), pages 7 à 41

<https://www.cairn.info/revue-economique-de-l-ocde-2002-1-page-7.htm> (consulté le 23/10/2019)

<sup>9</sup> Site de l'OCDE consultable en ligne :

<https://data.oecd.org/fr/> (consulté le 22/10/2019)

<sup>10</sup> Site de la Banque Mondiale consultable en ligne :

<https://donnees.banquemondiale.org/> (consulté le 22/10/2019)

<sup>11</sup> Notre base de données finale (au format CSV), *Google Drive* :

[https://docs.google.com/spreadsheets/d/e/2PACX-1vRs8JjiBAAeFhiQgClMbc6iMVn88lh14LjPz4jxAt eJNLQxnAy748VRHGuu\\_vBft1liMID\\_ARo1K6i/pub?output=xlsx](https://docs.google.com/spreadsheets/d/e/2PACX-1vRs8JjiBAAeFhiQgClMbc6iMVn88lh14LjPz4jxAt eJNLQxnAy748VRHGuu_vBft1liMID_ARo1K6i/pub?output=xlsx)

<sup>12</sup> Blampain.N, “L’espérance de vie par niveau de vie : chez les hommes, 13 ans d’écart entre les plus aisés et les plus modestes”, *Division Enquêtes et études démographiques, Insee*, 2018

<https://www.insee.fr/fr/statistiques/3319895> (consulté le 22/10/2019)

<sup>13,24</sup> Site du Café Pédagogique consultable en ligne :

<http://www.cafepedagogique.net/lemensuel/lenseignant/schumaines/ses/Pages/2008/Sant%C3%A9etPauvret%C3%A9danslespaysend%C3%A9veloppement.aspx> (consulté le 22/10/2019)

<sup>14</sup> Site Wikipédia, “Produit Intérieur Brut par habitant”, consultable en ligne

[https://fr.wikipedia.org/wiki/Produit\\_int%C3%A9rieur\\_brut\\_par\\_habitant](https://fr.wikipedia.org/wiki/Produit_int%C3%A9rieur_brut_par_habitant) (consulté le 08/10/2019)

<sup>15</sup> Droz-Georget.C “PIB par habitant : quelle tendance en Europe et en France ?”, *Office fédéral de la statistique*, avril 2019

<https://www.ifrap.org/etat-et-collectivites/pib-par-habitant-quelle-tendance-en-europe-et-en-france>  
(consulté le 09/10/2019)

<sup>16</sup> Site de l'OCDE, définition salaire moyen, consultable en ligne :  
<https://data.oecd.org/fr/earnwage/salaires-moyens.htm> (consulté le 15/10/2019)

<sup>17</sup> Site de l'OCDE, définition inégalité de revenus, consultable en ligne :  
<https://data.oecd.org/fr/inequality/inegalite-de-revenu.htm> (consulté le 13/10/2019)

<sup>18</sup> Site Wikipédia, définition indice de corruption consultable en ligne :  
[https://fr.wikipedia.org/wiki/Indice\\_de\\_perception\\_de\\_la\\_corruption](https://fr.wikipedia.org/wiki/Indice_de_perception_de_la_corruption) (consulté le 16/10/2019)

<sup>19,20</sup> Site de SES.WEBCLASS, "indice de développement humain IDH", *Chapitre 1. Quelles sont les sources de la croissance économique ?*, SES.WEBCLASS consultable en ligne :  
<http://ses.webclass.fr/notion/indice-developpement-humain-idh> (consulté le 17/10/2019)

<sup>21</sup> Jahan.S, "Rapports sur le développement humain", publié par *le Programme des Nations Unies pour le développement* (PNUD), 2017  
<https://www.undp.org/content/undp/fr/home/librarypage/hdr/> (consulté le 15/10/2019)

<sup>22</sup> Site de l'INSEE, définition du taux de chômage, consultable en ligne :  
<https://www.insee.fr/fr/metadonnees/definition/c1129> (consulté le 17/10/2019)

<sup>23</sup> Clément.M "Mieux comprendre les facteurs de risque de pauvreté en conditions de vie en contrôlant les caractéristiques inobservées fixes", *Insee – DSDS – Unité des études démographiques et sociales*, p.57 (consulté le 20/10)  
<https://www.insee.fr/statistiques/fichier>

<sup>25</sup> "Dépense intérieure de recherche et développement / DIRD", *INSEE*, 2019.  
<https://www.insee.fr/fr/metadonnees/definition/c1226> (consulté le 28/10/2019)

<sup>26</sup> "Dépenses extérieures de recherche et développement des entreprises (DERDE)", *Enseignement supérieur - recherche*, 2002.  
<https://www.enseignementsup-recherche.gouv.fr/reperes/telechar/res/res00/rapmeto00.pdf>  
(consulté le 28/10/2019)

MERCI !